Wireless Plug and Play Control Systems: Hardware, Networks, and Protocols

Meybodi, Soroush Afkhami

Publication date:
2012

Document Version
Early version, also known as pre-print

Soroush Afkhami Meybodi

Wireless Plug and Play Control Systems: Hardware, Networks, and Protocols

Wireless Plug and Play Control Systems: Hardware, Networks, and Protocols
Ph.D. thesis

ISBN: xxx-xx-xxxxx-xx-x
March 2012

Contents

Preface

This thesis, in form of a collection of papers, is submitted in partial fulfillment of the requirements for obtaining the Doctor of Philosophy degree in Automation & Control, at the Department of Electronic Systems, Aalborg University, Denmark.

The work has been carried out from November 2008 to December 2011, under supervision of Associate Professor Jan Dimon Bendtsen and Associate Professor Jens Frederik Dalsgaard Nielsen.

The PhD project was defined as one of the work packages of an extensive university project called *Plug and Play Process Control (P^3C)*. P^3C was partly sponsored by the Danish Research Council for Technology and Production Science. It was carried out as cooperation between Skov A/S, Dong Energy, Danfoss A/S, Grundfos A/S, FLSmidth Automation, The University of Rome La Sapienza, and Aalborg University.

Acknowledgements

I would like to express my gratitude to everybody who participates in making an enjoyable and peaceful working environment at Aalborg University, Section of Automation and Control. Among them, I should thank my advisors, Dr. Jan Dimon Bendtsen and Dr. Jens F. Dalsgaard Nielsen who stood patiently beside me in this period.

I would also like to thank Dr. Mischa Dohler, a distinct scientist, who gave me the opprtunity to stay at *Centre Tecnològic de Telecomunicacions de Catalunya* for five months and build the foundation of my work there.

Nowadays, friendships are not affected by long separating distances like before. I have stayed in touch with my good old friends who are very knowledgeable and influential in the institutes they are working at, across the world. I especially would like to thank Liu Arastoo, Shang Malek-Es, and Samandar Askarpour for sharing their great ideas on various topics and for the wonderful times that we have had during these years.

I am, and always have been deeply indebted to my parents, Nahid and Jalal, who took the responsibility of having and raising me during the most difficult years that my home country has gone through. They did it the best way I can imagine.

There is someone whose name should have appeared as a co-author in all of my papers. She is my closest colleague with whom I had many constructive discussions, my kindest friend who supports and encourages me every day, and my *genuine-companion*, literally translated from the Danish word: *ægtefælle*. Dear Maryam, I wish you health, hope, and happiness.

Soroush Afkhami Meybodi
March 2012, Aalborg, Denmark

Abstract

This dissertation reports the result of efforts to identify and solve the problems that arise when a control system is to be designed for various industrial case studies of the Plug and Play Process Control (P^3C) project that require autonomous addition/removal of sensors, actuators and subsystems. The scope is, however, constrained to the challenges that are relevant to the wireless communication framework which provides underlying services for plug and play control systems. Other aspects of plug and play control systems are studied in other research projects within P^3C.

The main results of this PhD project are presented in two distinct areas which are: 1) Signal propagation in underground and confined areas, and 2) Access and Networking protocols that accommodate the required flexibility, scalability, and quality of services for plug and play control systems.

The first category finds application in only one of the P^3C case studies where all of the nodes of the wireless networked control system are placed underground and should be able to transmit data among themselves. It is not a trivial problem because the well known radio frequency electromagnetic waves face serious difficulties penetrating the damp soil medium. To overcome the challenge, all potentially useful signal propagation methods are surveyed either by reviewing the open literature, or by doing simulations, or even running experiments. At the end, Magnetic Induction (MI) is chosen as the winning candidate. New findings are achieved in antenna design of magneto-inductive communication systems. They are verified by simulations and experiments. It is shown, via simulations, that MI is a reliable signal propagation technique for the full-scale case study: *Distributed Control of the New Generation of District Heating Systems*. The experimental results second simulation outcomes, but only in small scale. Therefore, it is essential to run full-scale tests in order to verify viability of the solution for the real problem.

The second category is an indispensible part of any network-based application, and is studied in sake of selecting the protocols that can fulfill the stringent requirements of P^3C case studies in general. In this part, after a thorough review of available Access and Routing protocols in industrial wireless networks, a combination of a deterministic Medium Access Control (MAC) protocol and a clustered flooding-based routing is recommended for wireless plug and play control systems. Formation and maintenance of clusters of nodes are directly linked to the top level application layer via a novel application-based routing metric. The proposed routing metric facilitates implementation of the networking topology in accordance with the control topology. Functionality of the new routing metric is verified by simulations within a flooding-based routing algorithm.

The results of this research project create a clear and concise interface for the other work packages of P^3C which are concluded earlier.

Synopsis

Denne afhandling rapporterer resultatet af bestræbelserne på at identificere og løse problemer, der opstår, når et styresystem er at være designet til forskellige industrielle casestudier af Plug and Play Process Control (P^3C) projekt, der kræver autonom tilføjelse / fjernelse af sensorer, aktuatorer og undersystemer. Omfanget er dog begrænset til de udfordringer, der er relevante for den trådløse kommunikation ramme, som giver underliggende tjenester til plug and play-kontrolsystemer. Andre aspekter af plug and play kontrolsystemer er undersøgt i andre forskningsprojekter.

De vigtigste resultater af dette ph.d.-projekt præsenteres i to forskellige områder, som er: 1) Signal formering i underjordiske og begrænset områder, og 2) adgang og netværksprotokoller, der tilgodeser den nødvendige fleksibilitet, skalerbarhed, og kvaliteten af tjenester til plug and play-kontrol systemer.

Den første kategori finder anvendelse i kun et af P^3C casestudier hvor alle knuderne af den trådløse netværk styresystemet er placeret i jorden og bør være i stand til at overføre data imellem sig. Det er ikke et ubetydeligt problem, fordi de kendte højfrekvente elektromagnetiske bølger stå over for alvorlige vanskeligheder trænger den fugtige jord medium. For at overvinde udfordringen, er alle potentielt nyttigt signal formeringsmetoder undersøgt enten ved at gennemgå den tilgængelige litteratur, eller ved at foretage simuleringer, eller endda kører eksperimenter. Ved slutningen er magnetisk induktion (MI) vælges som vindende kandidat. Nye resultater er opnået i antenne design af magnetisk-induktive kommunikationssystemer. De er verificeret af simuleringer og eksperimenter. Det er vist, via simuleringer, at MI er en pålidelig signaludbredelse teknik til fuld skala case study: *distribueret styring af den nye generation af fjernvarmesystemer*. De eksperimentelle resultater bakkes af simulation resultater, men kun i lille målestok. Derfor er det nødvendigt at køre fuldskalaforsøg for at kontrollere levedygtighed af opløsningen til den virkelige problemer.

Den anden kategori er en uundværlig del af ethvert netværk-baseret program, og bliver undersøgt i skyld at vælge de protokoller, der kan opfylde de strenge krav i P^3C case studier i almindelighed. I denne del, efter en grundig gennemgang af tilgængelig Access og routing protokoller i industrielle trådløse netværk er en kombination af en deterministisk Medium Access Control (MAC)-protokollen og en grupperet oversvømmelse-based routing anbefales til trådløs plug and play-kontrolsystemer. Dannelse og vedligeholdelse af klynger af noder er direkte forbundet til det øverste niveau applikationslaget via en ny anvendelse-baseret routing metrisk. Den foreslåede routing metriske letter implementeringen af netværk topologi i overensstemmelse med kontrollen topologi. Funktionalitet af det nye routing metriske er blevet bekræftet af simuleringer inden for en oversvømmelse-based routing algoritme. Resultaterne af dette forskningsprojekt skabe en klar og præcis interface til de andre arbejdssteder pakker af P^3C, som er indgået tidligere.

1 | Introduction

There are increasing interests in more flexible, scalable and self managing control systems. The project Plug and Play Process Control (P^3C), among the other academic approaches, was aimed for another step forward towards such systems. This thesis is meant to cover Work Package 1 (WP1) of P^3C. It deals with the problems that arise when designing wireless communication networks for plug and play control systems.

1.1 Motivation

The term *Plug and Play* (PnP) is well established when it comes to computer accessories. It basically refers to a device with specific known functionalities that could be utilized as soon as it is attached to a computer, with the least required effort to make it installed and running. The main idea behind PnP control is to extend the concept of PnP from computer accessories to the building blocks of a control system which are sensors, actuators and sub-systems with more general functionalities.

The variety of applications of control system devices makes their employment not as straightforward as computer accessories. For instance, when a USB flash memory is attached to a computer, the operating system makes it ready to be read/written as a temporary mass storage device. But, in a similar scenario, what should a control system do if a new temperature sensor pops up and starts reporting temperature of a vessel in a large process to the controller unit? Which of the actuators will be affected by the data from the new sensor? And in case of multiple controller units distributed in the network, which of them should make / keep communication links to the newly added device? These are some of the questions that reveal how complicated the challenges in building a PnP control system are.

Normally, designing control logics and algorithms is the last step in the system setup, after a model is developed for the existing system with known sensors and actuators. Making sensors and actuators plug and play, requires the controller to be able to identify and cope with changes in the system model and sensor-actuator pairing. Building a PnP control system involves several tasks, namely:

1. Automatically making communication links between added devices/units and the existing system through an integrated self-organizing self-healing communication network
2. Identifying types/locations of instruments that have joined the system, and finding what kind of measurements/actuations they offer.

3. Automatically updating the system model which is stored in the controller(s), if model-based control is of interest.

4. Automatically updating control algorithm(s) to maintain/improve control performance.

Combination of above mentioned tasks construct all of the six WPs of P^{3}C [P3C06]. This thesis focuses on the first WP, i.e. the required communication framework. Other work packages are covered in other PhD or research projects.

Regarding realization, WP1 is the most infrastructural part of P^{3}C because its goal is to design a flexible network of devices which can communicate in a reliable, scalable and secure manner. It means that the communication network must tolerate removal (deliberately or by faults) and addition of devices from different vendors without interruption, and interference with other existing systems. The role of WP1 is more explained in the following example.

Motivational Example: Fig. 1.1 illustrates a specific motivating example where a multi-input multi-output plant is considered with several sensors and four actuators. Four multi-input single-output control loops are implemented in three controller nodes in a wireless network.

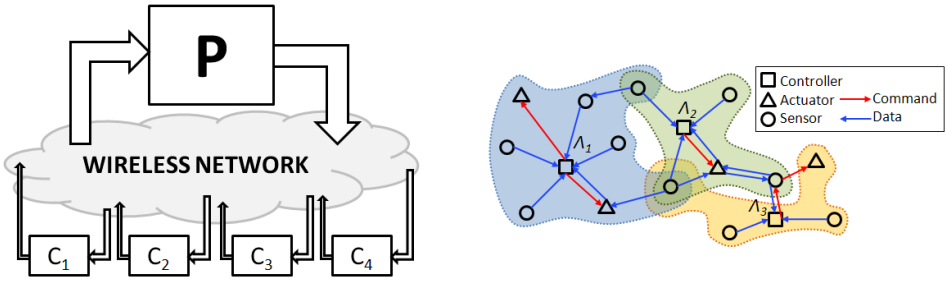

Figure 1.1: Block diagram of a networked control system with four multi-input single-output controllers (left) and its physical layout implementation in a wireless network with three controller nodes (right)

In this example, in case of addition of a new sensor, the role of WP1 is to route the data of the new sensor to all three controller nodes for a sufficiently long time interval, until the controller nodes can decide whether to stay connected to or dismiss the new sensor based on its usefulness for their control objectives.

In case of addition of a new actuator, WP1 should provide solutions such that all three controller nodes will be informed about presence of the new actuator and can quantify their own influence on the added actuator. Gradually, one of the controller nodes should succeed in taking control of the new actuator and should stay connected to it. All of the other controller nodes should let the new actuator go.

The above example shows the scope of WP1 and clarifies its interface to WP2 in which new measures are developed to quantify usefulness of the added sensors and actuators.

P^3C is motivated and financially supported by several industrial case studies which include:

- PnP climate control of a livestock stable

  The company that has brought this case study is seeking a control system that can be scaled up without human intervention; a control system capable of discovering and exploiting plugged-in temperature sensors at every corner of a large stable to improve climate control.

- PnP control of refrigerators in a supermarket

  A modular control system is sought in this case study, such that a bank of compressors could be efficiently utilized when the number and placement of fridges in a supermarket change.

- PnP control of a district heating system

  In this case study, a vast urban-wide distributed control system is of interest for a district heating system. The motivation is to have less heat loss from pipes' walls and more energy-efficient pumping. The idea is to implement a new piping layout which fulfills both of these objectives. Such a piping layout requires automatic addition/removal of new heat loads, i.e. newly built buildings, to the district heating network. Hence, this case study demands a PnP control system.

Based on the initial definition of this PhD project and the emphasis of industry representatives, a *wireless* network is to be used in the P^3C project. More flexibility in physical location of envisioned sensors/actuators and no idle cabling for probable future extensions support this choice. Moreover, the intriguing benefits of eliminating cables feed more stamina into the research problem of designing an industrial wireless framework. In what follows, it is shown that what kind of problems in an industrial wireless network may arise and which of them are of particular interest in this project.

1.2 The ISO/OSI model

Before highlighting the specialties of industrial wireless networks, it is beneficial to introduce the universally adopted standard for the structure of a networked application. It is known as Open System Interconnection (OSI) model which is developed by the International Standards Organization (ISO). It is also called ISO/OSI model. The overview of a networked application with two agents is shown in the OSI framework in Fig. 1.2.

Not every networked application complies with all details of this standardized framework. However, the OSI model introduces all of the concepts that might be of interest in any kind of networked application.

As an example, a simple data transfer between two agents in a fully OSI compliant network is described as follows: Suppose that the *Application* program of agent A has some data that have to be passed to agent B. The data must be formed in a machine independent format to be able to be presented to any other agent. This job plus potential data encryption is done in the *Presentation* layer. Establishing a network *Session* is the next step which is done by a transmission control protocol. It also controls and terminates the session when required. Keeping track of success/failure of data *Transport* is the

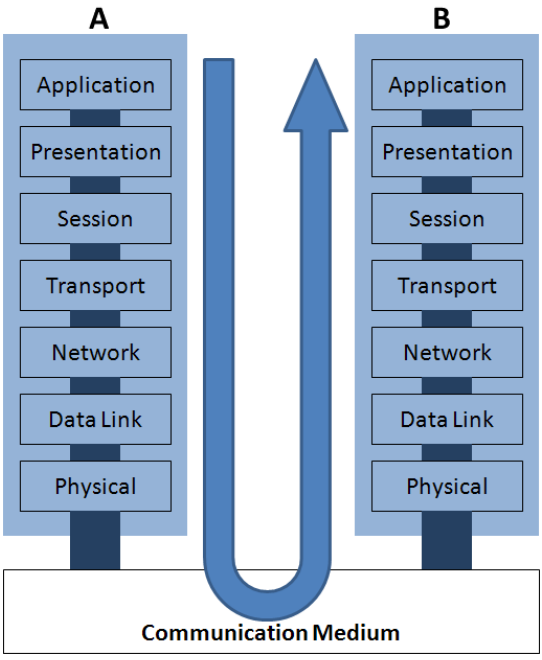

Figure 1.2: The OSI model of a networked application with two agents. The U-shape arrow describes a simple data transfer from node A to node B.

duty of the data transfer protocol which also decides if retransmission of data is needed in case of a failed transmission attempt. After being registered at the transport layer, a data packet needs to be informed about the route to its destination. This is done in the *Network* layer by a routing protocol where typically a sequence of addresses of relaying points plus the destination address are attached to the data packet. From now on, the packet stays alert on the *Data Link* protocol to get permission to be sent to the communication medium. As soon as the permission is granted, the packet is streamed into a *Physical* circuit which transforms its bits into standard symbols that are modulated and transmitted to the medium. Every other circuit that is attached to the medium and can hear the symbols demodulates them into a data packet. From the application layer on top of the stack of agent A, each layer typically adds some additional protocol dependent data, called *overhead*, to the original application data. On the recipient agents, these portions of data are detached from the original data step by step in a reverse order, as the packet moves up to the application layer. If a data packet survives this long path, it is received at the destination agent B.

The ISO/OSI model is applicable to any networked application irrespective of the protocols that are implemented or the communication medium that is used. However, some implementations have omitted some OSI layers by abstracting their functionalities in a single protocol. This is specifically true for higher layers in the stack. Nonetheless, in some cases, protocols in lower layers are also combined. Such designs are called *cross-layer* designs.

1.3 Wireless Networks in Automation Industries

The field of industrial automation has witnessed several quantum leaps in the course of time. The most recent one is the advent of industrial wireless networks which are brought to the field by the Highway Addressable Remote Transducer (HART) foundation, that is an active standardization community in industrial networking . Their latest specification in 2007, i.e. HART7, includes WirelessHART. Standardization movements have been continued by the International Society of Automation (ISA) at ISA100 committee in sake of providing a unified approach in wireless networking in automation industries.

The motivation behind, is quite appealing: No wiring, i.e. no cable conduits or trays among field instruments and control rooms, anymore. This means saving money in bulk material and in engineering, plus providing more available space and a tidier setup at the field floor. It reminisces the magnificent invention of digital fieldbuses that realized the idea of reliably using a single cable to carry multiple signals. They have offered similar benefits as mentioned above. Some examples are Profibus, DeviceNet, Foundation Fieldbus, etc.

The successful experience of industrial buses had increased expectations, but wireless turned out not to be a viable replacement for wire. Gradually, it was accepted that a different scope and application domain with milder timing requirements should be assigned to industrial wireless networks [NTWb]. It is constrained to *equipment monitoring* and *asset tracking*, in addition to *open loop control* in which control loops are closed by an operator. This strategy considers wireless as a complement, not a replacement, for wireline networks in order to extend their functionality.

In spite of the above argument, P^3C looks into applications with tougher timing requirements. It investigates feasibility of *closed loop control* in Wireless Networked Control Systems (WNCS) which are subject to topological modifications and even harsh signal propagation environments, in some cases.

1.4 Inherent Problems in Industrial Wireless Networks

Wireless networks' lack of performance compared to wireline networks, is not due to immaturity of the technology. It is because of the extra problems that should be dealt with in a wireless framework. These problems are inherent and always do exist. How we manage to deal with them defines how close the performance of industrial wireless networks could get to their wireline counterparts.

Apart from network security issues which attract experts from other fields of profession, there are two main particular challenges in industrial wireless networks which are: 1) stochastic nature of the quality of wireless links that affects reliability; and 2) additional end-to-end delay due to the limited coverage range of wireless nodes which urges a multi-hop communication framework in many cases.

In wireline networks, the first problem is not seen at all, provided that the electromagnetic compatibility (EMC) considerations are taken into account; the channel is a solid copper wire/bus which is laid in an accurately designed way and has time invariant characteristics. The second problem is seen only in large wireline networks where incorporating repeaters and bridges are inevitable among different segments of the network. In such cases, relaying messages between different segments imposes an additional fairly

fixed delay. Nonetheless, both of the mentioned problems are more serious and always exist in a multi-hop wireless network.

1.4.1 Lack of Reliability

Reliability is of utmost importance in industrial networks. It is defined as the fraction of sum of healthy packets that are received at all nodes over sum of transmitted packets from all nodes. There is no failed transmission in a 100% reliable network. Lack of reliability in wireline networks is mainly due to collisions at the MAC layer. In wireless networks, however, there is an additional reason for that which is the variation in quality of wireless links.

Quality of wireless links are prone to degrade due to: 1) Transmission power variations as a result of depleting batteries; 2) propagation losses if the transmission medium absorbs energy of radio frequency electromagnetic waves; and 3) propagation losses variations because of moving objects which may block the signal path or impose refractions/fading eventually after the wireless network is set up.

Among the above mentioned sources of quality loss, the second one, i.e. propagation losses in absorbing media, is the main concern in this thesis. It is associated with the control of district heating systems as one of the P^3C case studies. In that case study, nodes of the wireless network are confined underground where propagation losses in damp soil are significant. Therefore, signal propagation in underground and confined areas is the first problem that is investigated in this thesis in details.

1.4.2 Additional End-to-End Latency

From standing point of a control system designer, every networked control system has to be aware of communication constraints. They affect the control system by imposing an upper limit on the amount of data that can be transferred from a source to a destination with a minimum delay. Such a delay is called *latency* and the amount of data is known as the maximum *bandwidth* that the network can accommodate.

Given specific communication constraints, control system designer tries to minimize the required amount of data to be transmitted by aggregation, quantization and similar signal processing techniques, such that full bandwidth utilization results in acceptable bounds on the maximum end-to-end (e2e) latency.

It does not matter if it is a wireline or wireless network. Limitations on bandwidth and latency relate to the signal propagation method, PHY parameters and the MAC protocol. However, in multi-hop wireless systems, there exists another source of e2e latency which is not deterministic and introduces variations in latency, i.e. *jitter*. It occurs in multi-segment wireline and multi-hop wireless networks where a packet needs to be relayed via a number of intermediate nodes to arrive at its final destination. This type of latency is examined in this thesis by reviewing various routing protocols in multi-hop wireless networks followed by proposing a control oriented routing algorithm. Fig. 1.3 highlights the problems that are selected and will be tackled in this thesis.

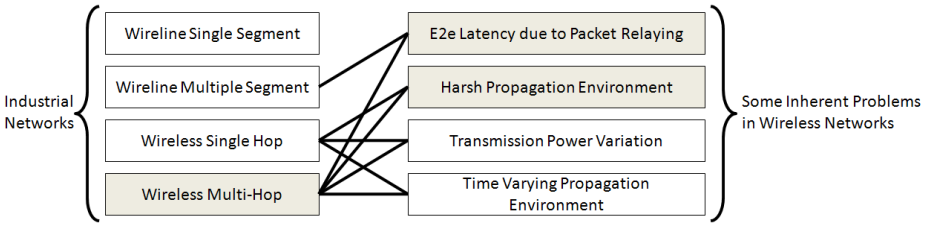

Figure 1.3: Links between industrial network types and some challenges - Selected network type and challenges in this thesis are grayed.

1.5 Research Questions of the PhD Project

The goal of this PhD project was to offer a solution for at least all of the P^{3}C case studies with emphasis on more general approaches and solutions, such that they can be directly applied to any other similar system. To achieve this goal, a questionnaire was distributed at the beginning of the project among the industrial partners of P^{3}C, each of which representing one of the case studies. Tab. 1.1 shows the discussion results with those people. It indicates the variety of some decisive characteristics of the three considered case studies.

Table 1.1: Comparison of P^{3}C case studies based on networking related characteristics

Case Study	One or Many Controllers	Line or Battery Powered	Propagation Environment
District Heating	Many	Line	Underground
Supermarket Refrigerators	One	Line	indoor-NLOS*
Stable Climate	One	Battery	indoor-LOS**

* Non-Line of Sight
** Line of Sight

Referring to Tab. 1.1, *the number of controllers* has a direct influence on the networking topology. While centralized control system structure is compatible with single sink converge-cast sensory data traffic pattern, a decentralized or distributed control system structure suggests an ad-hoc peer-to-peer networking topology. Knowing the networking topology is the first step to select/design compatible access and networking protocols.

Constraints on *power consumption* is specifically important in MAC protocol design. If the nodes are supplied by unlimited power sources, minimizing e2e latency will become the sole design criterion. Otherwise, a compromise should be made between timeliness requirements of the application and the networks lifetimes.

Finally, *propagation environment* has a direct effect on feasibility of making communication links among the nodes. The PnP control of district heating system case study added a major workload to the project that must have been dealt with. That was the only case study with a different propagation environment, namely damp soil. Electromagnetic waves in radio frequency range face serious propagation losses in damp soil which adversely affects coverage range of individual nodes.

Based on the above facts, two research questions were identified for the PhD project and the work in WP1 was carried out in two distinct phases in order to answer each

research question separately.

1. **What is the most viable method to create communication links among nodes of an underground sensor network whose neighbors are located several tens of meters away from each other?**

2. **What is the best access and networking protocols for reconfigurable control systems that can be either centralized with constrained power or distributed with unconstrained power sources?**

1.6 Outline of the Thesis

Having defined the domains of interest, it is time to outline the organization of the rest of the thesis. The next chapter offers state of the art and background on solutions and protocols of wireless industrial networks within three lower layers of the OSI model.

The methodology of conducting this research project is explained in Chapter 3. From the beginning, it is divided into two section, each of which addressing one of the research questions of the project. The first section addresses physical communications and the second one studies networks and protocols.

For better visibility and readability, Chapter 4 enumerates the contributions of the thesis in a bullet-point style list. It is indicated in this short chapter, what appendices should be looked into to find about a specific contribution.

The thesis is wrapped in Chapter 5 where conclusions and possible future works are stated. Finally, a list of references followed by all of the prepared research articles during this project is enclosed to complete the thesis.

Last but not least, the body of the thesis is written in a way that the reader can read it seamlessly from the beginning to the end without really requiring interrupting and referring to appendices for further details. This is provided to facilitate reading of the thesis and has inevitably forced the author to replicate blocks of text and figures from the articles in the body of the thesis.

2 State of the Art and Background

This chapter consists of three sections, each one devoted to review the literature on solutions and designs in one of the lower ISO model layers. As highlighted in the last chapter, only the relevant literature for industrial wireless multi-hop networks is studied.

2.1 Physical Communications in Underground and Confined Areas

The motivation to study signal propagation in underground and confined areas originates from the challenging signal propagation conditions of one of the P^3C case studies, i.e. control of the new generation of district heating systems.

District heating is a method for providing heat demands to closely located buildings by delivering hot water to them and cycling the return cooled water into heating plants via dedicated pair-wise pipelines. They create an urban-wide pressurized hydronic vessel in which water flow is made possible by centrifugal pumps. Pipes are buried underground and pumps are placed in the basement of buildings along the pipes. They are wired to local pressure sensors and can provide local control of the hydronic dynamics via their variable speed drives.

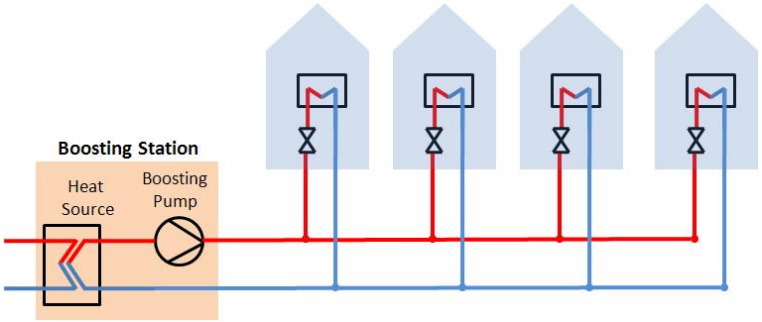

Figure 2.1: District heating piping layout: The traditional design

In the traditional design (Fig. 2.1), there are a few pumps for many end users. Assuming a minimum required pressure across all end users, each pump should be operated based on the required differential pressure across the farthest end user [DHS]. In old days,

a high enough differential pressure was being produced across each pump that to ensure the minimum required differential pressure across the farthest load at maximum flow. Today, the measured pressure across some test loads could be transmitted sporadically over cell phone infrastructure at prescribed time intervals. A human operator uses this data besides weather conditions and weather forecast, to roughly change the pump operating settings. A disadvantage of such control method is the considerable loss of pumping energy due to overpressuring across end users which are located close to the pump. The extra pressure needs to be compensated by involving pressure reducing valves at inlet of each end user. Furthermore, the diameters of pipes are chosen so large to prevent pressure loss in far loads. This increases heat loss due to large lateral surface, and hence deteriorates overall energy efficiency.

In a newly proposed piping layout [BVR+04] which is designed to improve pumping energy efficiency and alleviate heat loss, the number of pumps are significantly increased. In return, each pump provides a lower differential pressure, thereby eliminating very high pressure zones in the network. Moreover, cooperation among the pumps is mandatory in this design to maintain stability of the hydronic network (Fig. 2.2).

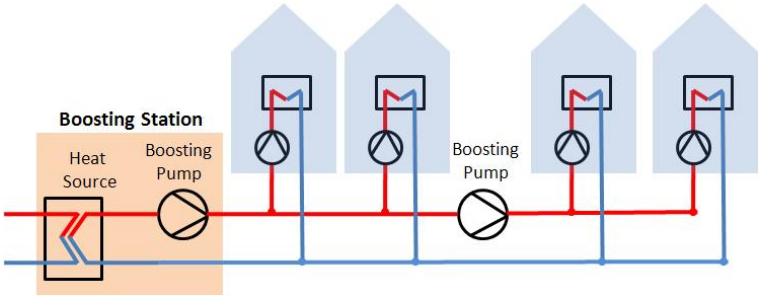

Figure 2.2: District heating piping layout: New design

The goal is to divide the burden of the big old pump among smaller new pumps in an optimal manner. The new pumps are to be placed at every end user and in boosting stations. This layout results in saving pumping energy and facilitates the use of pipes with smaller diameters in future installations which contributes to heat loss alleviation. On the other hand, this structure needs a more complex control system that requires data communication among pumps, e.g. data from end user pumps must be sent to their upstream boosting pump [Kal07, PK09].

Signaling among the pumps is not trivial. They are located in buildings' basements, which are typically air-filled underground chambers with concrete walls. Pipes are already laid and there is no chance to add wires along the current pipes to connect them together. Therefore, a wireless solution is mandatory.

2.1.1 Radio Frequency Electromagnetic Waves

Several methods have been tried in the literature to solve the underground signal propagation problem, but there is still no outstanding solution. Above ground, it is usually adequate to adopt any of the IEEE standards at the physical layer of wireless networks. They make use of certain intervals in the electromagnetic radio frequency spectrum which

are not reserved by governmental services. These intervals are known as Industrial, Scientific, and Medical (ISM) bands. Examples of applications include cordless phones, bluetooth devices, microwave ovens, wireless LAN, etc. It is basically legal to transmit in the ISM bands for personal, professional and amateur purposes. ISM bands are available in a wide range from several MHz to a couple of hundred GHz. How to choose the suitable ISM band depends on several factors including, propagation losses in the environment, exogenous interferences, and the required bandwidth by the application. While higher frequency EM waves potentially provide more bandwidth, they are more prone to absorption and refraction by the objects placed on the line of sight between transceivers. In other words, they require a cleaner environment for a predictable pattern of propagation.

As of underground problems, if using ubiquitous radio frequency electromagnetic waves in ISM bands were reliable enough, it would certainly be the best choice due to unprecedented availability and support of embedded radio chips. However, in underground systems, the main challenge in physical layer is coverage rather than interference mainly because of the high propagation losses of EM waves through the soil [AS06].

The maximum transmission range in RF systems depends on the following parameters:

- Transmission power
- Antenna gains
- Propagation losses
- Receiver sensitivity

Propagation losses in a homogenous environment consist merely of pathloss which depends on dielectric properties of the medium and distance between transmitter and the receiver. Dielectric properties of soil depend highly on the soil composition and water content [HUD+85, DUHER85] that varies quite a lot for different locations and from time to time. This results in a large uncertainty in pathloss behavior besides its typical large value.

Reference [ASV09] gives insight to the maximum achievable transmission range with different frequencies, transmission powers and volumetric water contents of soil. It is shown that the range can be extended only up to 4 m with 10^{-3} bit error rate (BER) for nodes buried in a depth of more than 2 m, assuming operating frequency 400 MHz, 30 dBm power transmission, volumetric water content of 5%, and PSK modulation which are all quite well selected parameters. However, this range is inadequate for our problem which requires a coverage range in order of 40-70 m [MPD10]. Moreover, higher water content and higher frequencies exacerbate the results.

Another approach is considering a combined underground above-ground scenario. The nodes in our network are not buried, but placed in the building basements. This will provide a chance for creating multiple paths for a signal, through walls from the basement to the surface, transmission above the surface, and through walls to the basement again. In an unrealistic ideal condition depicted in Fig. 2.3, i.e. with the existence of ideal reflecting surfaces above ground, the wireless signal should penetrate at least through four thick walls.

The arrangement in Fig. 2.3 is similar to the indoor room-to-room scenario in [KMH+08] with four thick walls. Pathloss through thick walls can be estimated based on the empir-

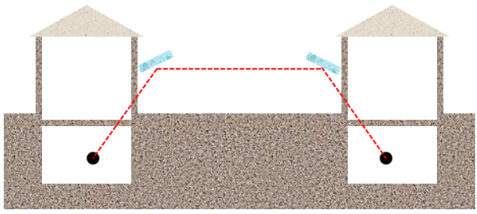

Figure 2.3: A typical underground above-ground combined scenario

ically derived pathloss models in [KMH+08]. The results are presented later in the next chapter. However, it turned out that the coverage range was not sufficient for our problem.

Another idea is to combine the ubiquitous embedded radios with the potential waveguide properties of the pipes. The question is whether a water filled pipe can act as an efficient waveguide. Although it has already been investigated with air filled metallic ducts with some degrees of success [TXS+04], there is a skepticism regarding applicability of water as the inner material. An EM waveguide principally consists of either two dielectric materials, i.e. a *dielectric waveguide* or a metal and an insulator, i.e. a *matallic waveguide*, but if none of the materials is an insulator, the wave cannot be propagated through the tube.

Water, as is used in district heating, is neither a good conductor, nor a good insulator. Although it has been subject to special treatments e.g. de-aeration and removing some ions and particles, its conductivity is in order of 10^{-3} S/m [DHS] compared to 2.0×10^{6} S/m for steel [NB85] and 0.3×10^{-14} to 0.8×10^{-14} S/m [wik10] for air. These numbers motivated us to do a simple experiment to measure signal attenuation in a water filled pipe presumed as a waveguide. Presentation of the results is postponed to the next chapter.

2.1.2 Acoustic Waves

Another intuitive option to make communication links among the pumps in Fig. 2.2 is to use the infrastructure itself as the communication medium. One way is to employ the mechanical vibration in form of sonic/ultrasonic waves as the carrier of data. Either pipes or water could be used as the main medium. There is an intrinsic advantage in using water because there could be variations in the material of the pipes in some cases in practice. Most of the times, pipes are made of steel and have welded fittings [DHS], but sometimes plastic pipes are used for smaller end user connections. Making use of pipes creates extra dependencies between the communication framework and the DHS, making both of them less flexible for future design modifications simply because the pipes are required to fulfill more design criteria.

Using water as the main communication medium has received less attention compared to in-pipe propagation. Sound speed in water is about 27% of in iron, i.e. 1500 m/s versus 5500 m/s [LCL03]. Therefore, the wavelength at the same frequency is shorter with the same proportion, making the wave more prone to destructive multipath effect. *Doppler Effect* is also another issue since the speed of water is comparable with the speed of sound [Sto08].

The most successful application in this area is underwater open-sea communication where the multipath effect is minimized. Usages are including oceanography, environ-

mental monitoring, *tsunami* warning, etc. [APM05] with commercial products.

Another working example which is more relevant to in-pipe communication is *mud pulse telemetry* [Kot88]. It is used to provide measurements while drilling oil and gas wells. While a pump at the surface pumps mud into the borehole and the pressurized mud drives the drill-head, the drill-head's built-in valves produce pressure pulses against the mud flow. The energy to create such pulses comes from the pressurized mud itself. These pulses are then received at the surface and the data are extracted. The typical commercial data rate is 12 bps for boreholes shallower than 7000 m. The highest reported data rate is about 40 bps for shallow boreholes down to 3 bps for boreholes deeper than 11000 m [WHD$^+$08].

Besides pipe's material and geometry, propagation of sound waves in a pipe depend on the materials inside and outside of it. Simplest results are achieved if there is vacuum inside and out. Simulation and experimental results are reported for gas filled pipes in [SNU06]. The application is remote reading of meters for natural gas distribution companies [SF06]. A two-way data transmission rate of 2560 bps is reported with 1×10^{-2} bit error rate (BER) for transmission via a typical six floor building with branches and bends in pipes, assuming in-line speakers and microphones. However, no commercial products were traced at the time of writing this thesis.

Wave propagation in water filled pipes is investigated in [LCL03], [Kon05], and [Kok06]. The second chapter of [Kon05] offers a clear and compact literature review on the subject, while [LCL03] gives a detailed analysis of the dispersion of different acoustic modes in a water filled pipe. Open-sea and in-pipe cases are compared in [Kok06]. All of these references state that the main source of signal attenuation in pipes is the reverberations which depend heavily on the geometry, fittings, bends and branches of the pipes that could be completely different from one case to another. It considerably reduces the achievable data rate by adding extra echoes, forcing us to use lower frequencies and requiring more sophisticated signal processing.

The most successful in-pipe application is pipe leak detector [Kok06] which is not a two way communication system and requires a low data rate. In such systems, microphones are installed on pipes with prescribed distances. They listen for a continuous noise which will be interpreted as the aftermath of water leakage from a pipe breakage. Provided that there is a synchronizing facility among microphones, the location of the breakage can be calculated by measuring the differences of time-of-arrivals of signals. There are commercialized cases on the market.

Experimental results [Kok06] for steel pipes with a diameter of 10 cm, predict the un-coded BER 32% for data sent at the rate of 7 kbps at 21 kHz with FSK modulation via water in a straight buried pipe. The transmitter and receiver were located approximately 9 m away from each other. Although the BER could be remedied by signal processing techniques, the transmission range is obviously insufficient for our requirements. Furthermore, the microphone and speaker which were used in this experimental study were installed inside the pipe and precisely centered at the pipe axis. However, in practice, clamped equipments are preferred to flanged or threaded ones due to feasibility and maintenance considerations.

From financial point of view, a general survey on commercial microphones, hydrophones and speakers show that there are some contact microphones and speakers that can be clamped on pipes, but they are not originally designed for in-pipe use. Therefore, the performance is not tested. Moreover, the price is comparable with a DHS small pump.

To get an understanding of the performance of the economical options, a simple test with fairly cheap microphones was carried out on a metallic pipe. Like the previous mentioned methods, the results for this experiment are also presented in the next chapter.

2.1.3 Power Line Communications

Power Line Communications (PLC) makes use of the power grid as a wireline network for data transfer. The idea has been around since 1920's for voice communication, telemetry, and supervisory control [SZ84]. The technology was revisited two times: 1)in late 1970's, motivated by the advent of digital signal processing, in order to be used as a means of load management and remote meter reading in electrical networks [Arr81]; and 2)in late 1990's to host broadband services like internet over power lines [Cla98].

The main drawback of using PLC is the variety of applications and non-interoperable standards which are thriving and competing to gain access to the readily available power grid infrastructure. Such a competition has become harsher than ever by the advent of Smart Grid. A comprehensive and recent study on applications and standards can be found in [GSW11]. Urban utilities, including electrical power distribution, water distribution, gas distribution, and district heating are not considered so far to use PLC in any application other than automatic metering. Therefore, finding space for bandwidth demanding applications like real time control of such plants involves lots of lobbying and non-technical settlements.

There are some technical difficulties in relying on PLC too. Frequency ranges in PLC start from a few hundred hertz for long distance narrow-band applications up to 200 MHz for home applications with a very short range and high bandwidth. The power grid is designed for 50-60 Hz which raises EMC considerations and mutual interference for employment of higher frequencies, especially for those more than 80 MHz due to the presence of ubiquitous TV broadcast channels. This fact has limited the permissible transmission power and has put off standardization processes.

Another problem is existence of transformers in power networks which filter out high frequencies. Therefore, bypass devices are needed to be installed parallel with transformers [Ols02]. This fact explains why PLC has been mostly successful in home automation and intra-building control systems where there is no electrical isolation; and low transmission powers can still cover the required range.

In a DHS, nodes are geographically dispersed and could be located in areas supplied by different transformers. Without the use of bypass devices, a partitioning pattern is imposed on the network which is probably not in agreement with the control system structure. If we accept this major limitation, PLC is a feasible low bandwidth solution, at least from the technical point of view, but maybe not a timely solution due to lack standards for our specific application and interference with other existing systems.

2.1.4 Pipes as Wires

This is another intuitive and cheap solution, provided that the pipes are all metallic and electrical conductivity is preserved across the fittings. The idea is to use the conductive pipes as electrical wires. There is only one major problem, i.e. Grounding.

In water distribution systems, pipes are in contact with soil either directly or through their thin layered coating. This forms a frequency dependent electrical resistance (R_{leak})

between the pipe and earth, as the electrical ground. Large R_{leak} is desired at the operating frequency. There is also another frequency dependent electrical resistance (R_{pipe}) between two neighboring nodes on the pipes. It is shown that with an appropriate use of the skin effect, by choosing the right operating frequency such that R_{leak}/R_{pipe} is maximized, it is possible to make use of this leaky multi-ground system to send and receive signals [EMS$^+$05].

However, in a DHS, pipes are heavily insulated with foam-filled PVC jackets in order to prevent heat loss. This may sound advantageous in the beginning. However, it makes it required that the pipes are grounded with copper wires as they enter each basement in order to hamper electrostatic shock hazards. Moreover, the following formula [NB85] shows that the skin depth of steel (pipes' material) is about 10 to 20% of copper (earth wire).

$$\delta = \sqrt{\frac{\rho}{\pi f \mu}} \tag{2.1}$$

In (2.1), δ is the skin depth, ρ the conductivity, f the frequency, and μ is the magnetic permeability of the conductor which varies significantly for different kinds of steel. Therefore, increasing the frequency will even aggravate the attenuation by increasing R_{pipe}/R_{leak}.

In summary, dependency on pipes material, direct grounding of pipes, and the stronger skin effect for pipes rather than earth wires altogether diminish any chance to transfer signals through the pipe by this method.

2.1.5 Cell Phone Infrastructure

There are lots of commercial products and some publications which claim to offer remote monitoring and control of home appliances by mobile phones. The idea is to use two SIM cards and transfer data by Short Message Service (SMS). It works well, when a single measurement or command needs to be sent. Switching home appliances ON and OFF, sending an alarm signal from the field to a remote operator, or automatic reading of electricity/water/heating meters once a day/month are among success stories, but it cannot be applied to our closed loop control problem due to the following concerns.

First of all, the system will be dependent on a third party infrastructure which is not desirable. Second, in every closed control loop, signaling should be fast enough compared to the dynamics of the measured variable, i.e. water pressure in our case. In hydronic systems, changes in water pressure propagate with the speed of sound in pipes, which is 1500 m/s. Comparing this speed with typical distances in an urban area suggests a sampling rate faster than 1 Hz. Sending a SMS at every second is quite expensive, if possible. Third, from technical point of view, SMS is sent via signaling paths which are originally provided to control the telephony traffic. They are separate from main voice path and have a maximum payload capacity of 1120 bits. The signaling paths let SMS go through whenever no higher priority traffic control signal exists. Therefore, congestion can happen resulting in a varying delay of several seconds. SMS centers store and retry to forward old messages by default. This function exacerbates the situation by making a long queue at the SMS center. Moreover, reordering of the gathered data may happen unless at least a local time stamp fills a portion of the payload and is stored in the receiver.

Furthermore, our application requires the nodes to be placed in the basement of buildings where signal strength is greatly attenuated even for well covered cellular networks. Based on my observations, even in automatic reading of meters which is a quite successful application for cellular monitoring, regular human operator check is required when the apparatus is located in the basement.

2.1.6 Magnetic Induction

MI has been around since 1920's being used as a means of communication between submarines [Bat20]. There is no fundamental difference between MI and RF EM waves. However, because of different technical specification, most importantly antenna size, available bandwidth and transmission power, various names have popped up to describe it including *Very Low Frequency (VLF)*, *Ultra Low Frequency (ULF)*, and *Extremely Low Frequency (ELF)*. Each of these names addresses a certain range of frequencies which are not even understood in the same way, in different communities. For example, *National Aeronautics and Space Administration (NASA)*, *World Health Organization (WHO)*, and *International Telecommunication Union (ITU)* have all assigned different frequency ranges to define the above terms. In this thesis, we use MI as the only alternative term and explain why it is worth being distinguished from usual radio spectrum.

MI also employs alternating Electric ($\mathbf{E}$) and Magnetic ($\mathbf{B}$) fields as data carrier, but it uses low frequencies and low power such that the communication range falls off within the distance predominated by the *static* or *quasi-static* fields, as described in the following.

$\mathbf{E}$ and $\mathbf{B}$ fields which are created by an alternating electrical current in an antenna can be estimated by the summation of several terms that have different dependencies to distance *r* from the antenna, e.g. $|\mathbf{B}| \approx \alpha.1/r^3 + \beta.1/r^2 + \gamma.1/r$ in which α, β, and γ depend on the antenna pattern and its driving current. In the vicinity of the antenna, fields are very complex (to be shown later in 3.2.2). Gradually, they are predominated by $1/r^3$ term. Up to this zone, is called the *static field* domain. Other terms predominate at farther distances and are called *inductance field* (*quasi-static field*) and *radiation field*, respectively [NB85]. The latter accounts for EM radiation in which $|\mathbf{B}| \propto 1/r$ and $|\mathbf{E}| \propto 1/r$. To define which field component is predominant, r should be compared with *wave number, k*:

$$k = \omega/v \tag{2.2}$$

where ω is the angular frequency of the $\mathbf{E}$ and $\mathbf{B}$ fields and v is the speed of EM waves in the medium. $k.r \gg 1$ and $k.r \ll 1$ imply that either the radiation or the static field is dominant, respectively. Otherwise, the field is quasi-static.

Static and quasi-static fields decay much faster than the radiation field while getting away from the current source. More concretely, their energy flux *Poynting vector* decays faster than $1/r^2$ and the emitted power per area tends to zero as distance from the source goes to ∞. Therefore, they are non-radiating fields and have no value in long range communications. However, if the frequency is low enough, the distances of interest can still be perceived as short distances. Hence, even non-radiating fields can be exploited to transfer energy between transceivers. More detailed studies on MI are carried out and the results are presented in the next chapter. It is shown that MI is a viable signal propagation method for our problem.

2.2 Medium Access Protocols of Industrial Wireless Networks

Wireless networks in automation and control domain are better known under the name *Wireless Sensor Networks (WSN)*. Unlike what this name suggests, WSNs may consist of any element in a control system including sensors, actuators, controllers, gateways, etc. In the last 15 years, a large number of MAC protocols have been devised particularly for WSNs. Before examining the most influential WSN MAC protocols in details, we review the classical taxonomy of them [BDWL10, WMW05, Wil08]:

- *Token-based MAC Protocols:* Token passing is originally designed for wireline networks. It is the chosen medium access control strategy for industrial wireline networks such as Profibus FMS and CAN bus. Thus, it would be very interesting if it is also applicable to industrial wireless networks. In Token passing MAC protocols, a single token is transferred among the nodes that may need to transmit data. Whoever has the token can take control of the channel and establish communications. After sending its data, the token holder node passes the token to one of its neighbors that are identified as its successor. This procedure continues to form a directional ring of token holders.

  Token passing method offers a deterministic behavior in controlling channel access. It guarantees Quality of Services (QoS) in terms of bounded latency and reserved bandwidth. It is realized by constraining the maximum number of nodes that can join a ring and the longest time interval that a node may possess the token before passing it to the next node.

- *Reservation-based MAC Protocols:* This category includes the protocols which provide a time schedule for every node, such that each one is aware when to access the channel and for how long keep control of the channel. Time Division Multiple Access (TDMA) is the representative of reservation-based protocols that guarantees a deterministic and collision-free channel access mechanism.

  Normally, the scheduler in reservation-based MACs requires knowing the complete and up-to-date topology of the entire network. Moreover, providing some sort of synchronization is essential to establish a common sense of time among all nodes. In summary, this family of protocols has demanding requirements and potentially offer ideal performance and optimal use of resources. However, they suffer from lack of scalability and reconfigurability.

- *Contention-based MAC Protocols:* This family of MAC protocols in contrast to reservation-based MACs is known for its simplicity which is attained by sacrificing performance. Nodes with contention-based MAC may have common active periods and are prone to collision occurrence at the PHY. The representative protocol of this family is Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) which includes a hand shaking mechanism to reduce probability of collisions.

  In comparison with reservation-based MACs, performance of contention based protocols is always inferior. However, the difference depends on traffic congestion. In unsaturated low load networks the measures of performance, e.g. throughput or reliability, are very close for both categories. As traffic increases, performance

of reservation based protocols remains constant until traffic saturation, but performance of contention based protocols begins to degrade and ends up in a jammed situation, well before the saturation point. That happens when no healthy transmission is practically possible and the retransmission requests are accumulated to make a virtual saturation.

- *Preamble Sampling MAC Protocols:* These protocols are specifically designed for nodes that have limited energy supply. In practice, many WSNs utilize battery operated nodes. Therefore, energy consumption has gradually become the main concern in designing WSN MAC protocols.

  Preamble sampling basically directs energy consumption to the nodes that are assumed to be line powered. It is not about reducing overall energy consumption. Instead, the goal is to redistribute the burden. For example, if there are a few line powered transmitting nodes while the receivers are battery powered, a long pulse is introduced in the beginning of each packet, namely *preamble*. This provides the chance for receivers to spend most of the time in sleep mode with the minimum power consumption. At specific intervals, they turn on their receivers and listen to the channel for a short time just to detect any ongoing preamble. Long preambles, in this case, save energy of receiving nodes, but result in increased power consumption in transmitting nodes and introduce considerable latency. When sacrificing performance is not acceptable, preamble sampling is not an acceptable approach.

- *Hybrid MAC Protocols:* Some newer protocols use a hybrid solution of reservation based and contention based MACs which requires a layered network topology. For instance, there could be different types of nodes with different levels of functionality. Some reduced function nodes may only contribute to simple star topology while the full function nodes could make a mesh network amongst themselves. In such a network, it is possible to assign different MAC principles for different types of nodes. More details and examples will be discussed shortly.

2.2.1 From Wireline to Wireless: Wireless Token Ring Protocol

Token passing is a distributed mechanism and does not require a base station to access all of the nodes in order to synchronize or schedule them. Therefore, it is a potential candidate for wireless networks with limited coverage range. However, it has been shown in [WW01] and [Wil02] that the performance of token passing MAC protocols over lossy wireless links may easily fall below the acceptable level. It is due to high sensitivity of the protocol to reliable transmission of *token transfer* messages. Temporary loss of the link that should pass the token to the successor node will result in exclusion of the successor from the token ring, and hence makes it isolated until it joins to the ring sometimes later. This is called the *ring stability problem*. While isolated, the node may not transmit any data irrespective of its importance. Therefore, it is not suitable to employ the MAC protocol of e.g. wired Profibus, in order to create the wireless version of this popular field bus.

Several researchers have tried to offer a wireless token passing protocol. A remarkable result is the Wireless Token Ring Protocol (WTRP) [ELSV04]. WTRP works in a distributed fashion like any other token passing protocol. It provides multiple channels

and spatial reuse in sake of capacity increase. It handles automatic topology reconfiguration when nodes join and leave rings. It came into attention mostly because of being the pioneer, being implemented in practice, being supported by an open source software simulator, and being extensively documented in two master theses and several papers [Lee01, Erg02]. The main contribution of WTRP is introducing several other control messages named after different kinds of tokens. Some of them deal with ring maintenance while the others handle automatic reconfiguration and scalability.

WTRP – Protocol Description: In WTRP, each node knows its predecessor and successor in a stable ring. The main contribution of WTRP is introduction of several other control messages named after different kinds of tokens. It actually facilitated automatic reconfiguration of rings and made self healing possible. These tokens are listed as follows.

- *Claim-Token*: Whenever a node creates a new ring, hence a new token, or regenerates a lost token after a ring recovery procedure, it broadcasts a claim-token. Claim-token acts as a normal token and only indicates the first occurrence.

- *Token-Deleted-Token*: Whenever a node receives a token, it examines the token to extract its *priority*. The priority concept is devised to prevent multiple tokens roaming in the same ring. Multiple tokens may be created due to multiple claim-tokens generated in different parts of a ring.

- *Solicit-Successor-Token*: If the token holding station decides to search for a new successor, either to expand the ring or to repair it after missing a successor due to a dead link, it broadcasts its token and passes the right to transmit to all of its 1-hop neighbors for a limited time. This broadcast token is called solicit-successor-token and temporarily renounces the right to transmit.

- *Set-Successor-Token*: When a station wants to join or leave a ring, first, it waits to get either a solicit-successor-token or the token. Then, it returns the right to transmit by transmitting a set-successor-token. For a planned leave, set-successor-token is transmitted as soon as the token is received and accepted at the station. However, for joining, the station waits for a random back-off timer after receiving solicit-successor-token and transmits set-successor-token before solicit-successor-token expires. The back-off timer provides a probabilistic collision avoidance solution among competing nodes in join process.

- *Set-Predecessor-Token*: Whenever a node selects a new successor, either to complete a *join* or *leave* process, it passes the right to transmit to the new successor by a set-predecessor-token.

All of the above mentioned tokens except solicit-successor-token unconditionally renounce the right to transmit. Claim-token and token-deleted-token deal with ring maintenance, while solicit-successor-token, set-successor-token and set-predecessor-token handle automatic reconfiguration and scalability.

Although WTRP offers several novel techniques to add scalability and reconfigurability to the traditional wired token ring protocol, the main concern still exists. WTRP gives proof for ring stability assuming no link failure. Although a ring recovery mechanism is

introduced, there is no evaluation of ring stability when the links are lossy. Thus, WTRP does not guarantee ring stability in presence of lossy links.

The only remedy proposed for token ring protocol with lossy links so far, is to use a *Polling based mechanism* to avoid token loss [Wil03]. Polling provides a rigidly deterministic channel access. A central node is responsible to poll all other nodes and ensure that the token is alive and frequently transmitted. Polling based MAC has proved to be robust in presence of uncertainties in wireless channel [Wil03]. However, every polling-based protocol or mechanism requires bi-directional links between a central station and every other node in the network which is impossible to provide in networks with nodes that have insufficient coverage range.

2.2.2 Designed for Industrial Wireless Networks: VANET & WPAN

Among the originally designed solutions for wireless networks, Wireless Personal Area Networks (WPANs) and Vehicular Ad hoc Networks (VANETs) are two distinct areas with agile standardization activities that have found grounds in various relevant applications.

WPAN, as the subject of the IEEE802.15.4 standard, focuses on multihop WSNs which consist mainly of battery-operated resource-limited nodes. WPANs have found many applications including home automation e.g. in Zigbee, industrial automation e.g. in WirelessHart and ISA100.11a, large scale environmental monitoring, asset tracking, etc. WPANs make use of reservation-based and hybrid protocols in the MAC layer with respect to their operation mode, to be explained shortly. On the other hand, VANETs, as introduced by the IEEE802.11p standard, refer to WSNs that consist of nodes that may splurge energy in order to keep up with the quickly changing topology and the tight e2e latency requirements of vehicular wireless networks.

In automation industry, the focus has been on WPAN so far, but it is understood that latency sensitive applications and closed loop control are far from being feasible in the current practice where the main concern is energy consumption. At the same time, high reconfigurability and tight e2e latency requirements of PnP control systems, bring VANETs into consideration. To further develop this initiative, the key properties of WPANs and VANTEs are summarized in the first two columns of Tab. 2.1, followed by the properties of an ideal network that hosts a PnP control system. We call it Controller Ad hoc Network (CANET) henceforth.

The similarities between VANETs and CANTEs in Tab. 2.1 suggest to eliminate the constraint on power consumption in CANET in order to provide a chance to satisfy requirements of latency sensitive applications. This means that the nodes will not be truly wireless since they have to keep their power cords. However, getting rid of signal cables is an advantage by itself. It is reasonable, at least for *actuators* including motors, pumps, and similar devices to always run on power cables and have dedicated wired infrastructure in *Motor Control Center (MCC)* in factories and industrial plants.

Tab. 2.1 also indicates that the nodes in CANETs have limited coverage range such that they cannot be synchronized by a central access point. Necessity of having a central coordinator is one of the basic constraints in most of the WPAN compliant protocols which considerably limits their applicability.

Table 2.1: Comparison of key properties in designing MAC protocols for wireless PnP control systems

Network Type	VANET	WPAN	CANET
Constrained Power	No	Yes	??
Latency Sensitive	Yes	No	Yes
Synchronization	*Yes (GPS)	*Yes (access point)	No
Central Coordination	*No	*Yes	No
Topology Change	Fast	Slow	Medium
Payload Volume	Small	Small	Small
Traffic Pattern	Ad hoc	Convergecast	Ad hoc

* Not Always

2.2.3 VANET in Automation Industry

While the IEEE802.11 defines PHY and MAC layer specifications for wireless LAN networks in general, the amendment of 802.11p specifies PHY and channel access mechanisms which suit vehicular environments. These include vehicle to road-side and inter vehicle communications whose key features, according to Tab. 2.1, are quite similar to CANETs.

802.11p enjoys seven channels at the PHY, designated to be used for different applications ranging from spreading time critical safety messages, e.g. vehicle collision warnings, to distributing supportive data, e.g. updated traffic information or even local commercial advertisements. Hence, all packets are marked with their priority, namely Access Category (AC), to be able to gain access to the appropriate channel. One of the channels is solely designed as the control channel and the other six could be used for data handling.

At the MAC layer, 802.11p has adopted the IEEE802.11e. 802.11e amends MAC enhancements to 802.11 in order to improve QoS. In general, 802.11e has two operating modes. One is based on polling nodes from a central station, hence is a contention-free mode and assigns guaranteed time slots to specific nodes. It is called *hybrid coordination function controlled channel access (HCCA)*. The other, which is the main operating regime in VANET, relies on distributed CSMA/CA and is called *enhanced distributed coordination access (EDCA)*.

The basic idea behind EDCA mechanism is to prioritize packets based on their data-content. It defines individual sets of channel access parameters such that high priority packets have a higher probabilistic chance to win the contention. The mentioned parameters are: 1) transmission delay after sensing the channel idle, and 2) the back-off time after unsuccessful transmissions. For example, when a high priority packet shall be transmitted, both of the parameters get smaller values. In summary, the EDCA mechanism in VANETs provides stochastic prioritization with adjustable parameters.

2.2.4 WPAN in Automation Industry

WPAN is known by the IEEE802.15.4 PHY and MAC layer specification for local and metropolitan area networks. Some examples are Zigbee, WirelessHART, and ISA100.11a.

Since ratification of the latest version of 802.15.4 in 2006, several amendments have been appeared to enhance its functionality in various directions among which activities in the *4e* working group deserves more scrutinization.

IEEE802.15.4e amends MAC enhancements to 802.15.4 for *industrial applications*. Its main purpose is to add two features of *high reliability* and *low latency* which are the most serious concerns in industry. 802.15.4e is in draft status yet, but some of its basic contents are mentioned in the sequel [WSL10].

Application domains in 802.15.4e are divided into two groups based on their latency requirement, namely the latency tolerant Process Automation (PA) and Commercial Applications (CM), and the latency sensitive Low Latency (LL) networks. LL specifications are so demanding that are only achievable by combination of the star topology, small payload, and TDMA deterministic channel access. For PA, on the other hand, a flexible design with multiple topologies, packet formats, and channel access mechanisms is feasible. In other words, a hybrid MAC protocol is envisioned for PA applications. The key aspect in PA, CM and LL cases is *determinism*; with the lowest possible deterministic latency for all LL, a tunable deterministic latency for PA and CM, and the highest feasible reliability, i.e. deterministic link quality for LL, PA and CM applications.

2.2.4.1 Deterministic Latency in the IEEE802.15.4e

In LL networks, a device could be either a sensor with merely uplink traffic or an actuator with bidirectional up/downlink traffic (Fig. 2.4). The gateway schedules TDMA access – one fixed time slot per node – for all of the nodes to create a superframe of timeslots. Superframe duration depends on the number of devices. It can be in order of 10 ms for a typical network of 20 devices. Superframes are synchronized by a beacon message transmitted frequently from the gateway.

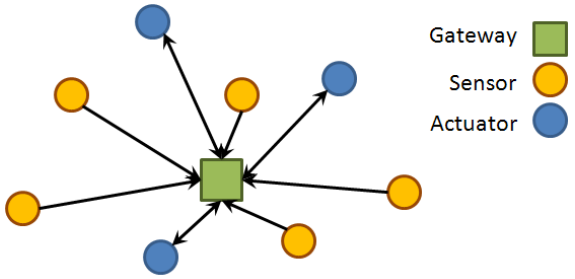

Figure 2.4: Network topology for low latency application

Special MAC frames with 1-octet MAC header are proposed to reduce MAC overhead, hence facilitating smaller PHY packets that fit into short timeslots and make superframes shorter.

Clearly, the use of multiple transceivers on different channels accommodates more devices in a superframe with specified length, but because there could be many LL networks in the proximity of each other, utilizing multiple channels is not allowed in sake of providing coexistence for LL networks.

For latency tolerant PA applications, a TSMP like protocol is proposed. Time Synchronized Mesh Protocol (TSMP) [PD08] is compliant with 802.15.4 and is the main ingredient of WirelessHART and ISA100.11a. Facing the problem of limited coverage range of nodes while requiring a deterministic cyclic performance forced TSMP inventors to consider superframes with fixed duration and lots of unused time slots which were reserved for accommodating the pre-defined maximum number of nodes that might join the network. This conservative approach makes performance of TSMP independent from the number of nodes by assuming the worst case. The length of the superframe defines e2e latency and is determined by the maximum number of nodes and an assumed physical distribution of nodes which approximately implies the number of re-transmissions among multiple hops.

In summary, if low latency is desired, 802.15.4e imposes star topology, hence requires a central access point that is inapplicable in wireless networks whose nodes have inadequate coverage range. In other words, if a flexible ad hoc topology is necessary, deterministic yet potentially high latency is the best achievable outcome.

2.2.4.2 Towards Reliable Wireless Links in 802.15.4e

In LL networks that assumingly take up small areas, reliability is provided by prudently selecting a single channel and constraining presence of sources of interference in the proximity.

For PA networks, an improved version of TSMP is utilized known as Time Slotted Channel Hopping (TSCH) [WMP09]. TSCH was the starting point in drafting 802.15.4e. This technique applies only to PA applications. The aim is to increase link reliability and robustness against: 1) time varying characteristics of wireless channel, and 2) temporary or permanent exogenous interferences.

In 802.15.4e, a link is defined by a time slot and a channel frequency offset. 802.15.4e makes use of 16 narrow-band channels in the interval 2.400-2.500 GHz commonly known as 2.4 GHz ISM band. Unlike *channel agility*, also called *channel adaptation*, in which the channel changes only after a problem occurs in the current channel, TSCH suggests a pre-planned channel hopping. In TSCH, an algorithm switches among different channels at each timeslot, with the ability to black-list some channels in sake of providing coexistence with other networks. Channel hopping and channel adaptation constitute *channel diversity* for 802.15.4e networks, which considerably increases probability of having interference-free wireless links.

Fig. 2.5 is reproduced from [WSL10] and shows the spectrum usage at 2.4 GHz ISM band in Europe and North America by three different 802 family standards. It includes 802.15.1 Bluetooth, 802.11 Wireless LAN, and 802.15.4e. The charts indicate that in both continents, 802.15.4e enjoys four interference-free channels. Therefore, if an environment-aware channel selective mechanism is in action, 802.15.4e will have an obvious advantage over Bluetooth and WLAN in terms of deterministic behavior.

Last but not least, channel hopping can also be carried out at the PHY layer. However, in PHY channel hopping, a packet is sent via a single channel even if it is a large packet and needs several timeslots. But in MAC channel hopping, the unit of transmission is a timeslot rather than a packet.

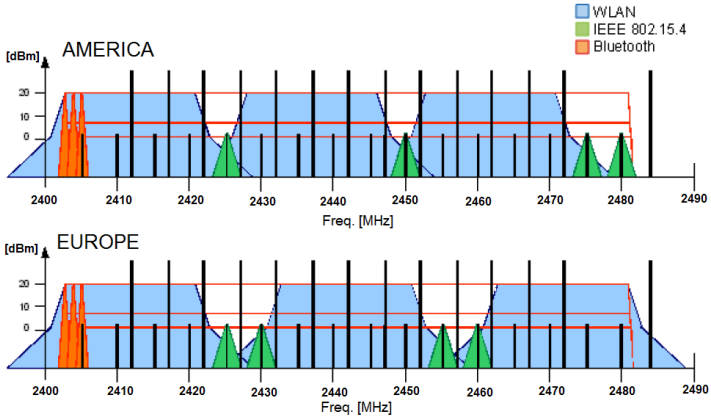

Figure 2.5: RF coexistence in the 2.4 GHz ISM band [WSL10]

2.2.4.3 Trade-off between Determinism and Energy Consumption

There are two optional Low Energy (LE) mechanisms in 802.15.4e. They work by defining a duty cycle for nodes such that energy consumption is compromised with QoS and the deterministic behavior. They are tentatively called Coordinated Sampled Listening (CSL) and Receiver Initiated Transmission (RIT).

CSL is useful when only the receiving nodes have limited energy resources. They sample the channel at low duty cycles to find out about any ongoing transmission It reminds of preamble sampling protocols. RIT, on the other hand, is appropriate when both transmitting and receiving parties have limited energy resources and high latency tolerance. In RIT, whenever a node needs to receive data, it initiates a data transmission by first transmitting a Request to Transmit message. Using this mechanism only makes sense if packets are transmitted infrequently compared to their required delivery latency.

2.2.5 VANET vs. WPAN

In summary, only PA application domain of the IEEE802.15.4e and the EDCA mechanism of the IEEE802.11e can be applied to PnP control systems. WTRP cannot appropriately handle the ring stability problem. Both LL networks of 802.15.4e and the HCCA mechanism of the IEEE802.11e require a central orchestrating node which cannot be provided when nodes have insufficient coverage range. Finally, RR-ALOHA was rejected due to large MAC overhead and other practical issues.

The bad news is that neither TSCH in the IEEE802.15.4e nor EDCA in the IEEE802.11e guarantee tight e2e latency requirements of PnP control systems. The former approach sacrifices the best achievable e2e latency by conservatively providing guarantees on latency bounds; and the latter cannot provide guarantees to its usually acceptable e2e latency. Therefore, TSCH in the IEEE802.15.4e is preferred when deterministic and fairly high latency is of interest. On the contrary, if low latency is crucial, EDCA in the IEEE802.11e should be employed with preparation to deal with jitter.

Last but not least, the IEEE802.15.4e has an obvious advantage over the IEEE802.11p when it comes to link reliability issues. Although the IEEE802.11p has 7 PHY channels, it does not employ them to provide a channel diversity mechanism as is done in the

IEEE802.15.4e. Instead, it exploits those channels to prioritize packets in sake of collision avoidance. Moreover, within each channel, it uses the EDCA mechanism to further prioritize packets. The IEEE802.15.4e is superior because it uses different resources for different tasks, i.e. different PHY channels to provide channel diversity, and an individual PHY channel for accurate scheduling.

2.2.6 Other Wireless MAC Protocols with Deterministic Performance

Self-Organizing Time Division Multiple Access (STDMA): is a reservation-based channel access mechanism which relies on GPS equipped nodes rather than a central coordinator [Lan96]. GPS time signal provides synchronization among the nodes and helps scheduling time slots, hence compensates lack of a central coordinator for providing deterministic channel access. Although STDMA is promising for VANETs with some commercial deployments, it is not suitable for CANETs where indoor affordable deployments are sought.

ADHOC-MAC: or equivalently *Reliable Reservation ALOHA* (RR-ALOHA) is originally developed for inter-vehicle communications [BCCF04]. It is a reservation based MAC protocol that relies on unconstrained power supply for each node in the network. It also considers limited coverage range of nodes and a true distributed mechanism in steady operation. It also does not rely on GPS or any similar central time synchronization method to facilitate TDMA channel access.

The basic idea in RR-ALOHA is that all of the nodes in a network are pre-programmed with the number of *time slots* in a *time frame*, and the duration of a time slot. When turned on, a node tries to reserve one of the time slots by chance. Collision can happen between competing nodes to reserve a specific time slot. When a node succeeds in reserving its own time slot, no other nodes in a two hop neighborhood may transmit during that interval. Therefore, this protocol completely eliminates *hidden terminal problem* and provides spatial reuse of the same time slot for far enough non-interfering neighbors. This is achieved in expense of obliging each node to gather information from its neighbors as far as two hops in order not to start any transmission during their reserved time slots. At the end, large MAC overhead imposed by the neighbor discovery messages significantly degrades performance of this protocol, especially in dense networks [BCCF02].

Another disadvantage of the ADHOC-MAC is in the initial phase of network formation, which is not a distributed process. It should start from a single point and grow from there node by node. It is not possible to start up the network at several positions and merge the segments when they overlap each other. Lack of time slot synchronization and potential conflicts in assigning time slots due to prior assignments in different segments are the main barriers for this purpose.

Last but not least, the effect of network density (X), as defined in (2.3), is not explicitly clear in selection of superframe length in ADHOC-MAC. Nevertheless, it has the same effect as it does in TSMP. That is, for denser networks longer superframes should be employed.

$$X = Max(x_j), \quad \forall j \in \{1...n\} \tag{2.3}$$

In (2.3), n is the total number of nodes in the network, and x_j is the number of nodes in a 2-hop neighborhood of the node j. With a constant n, higher density yields longer

superframes, and hence prolonged e2e latency.

2.3 Routing Protocols of Industrial Wireless Networks

In wireline industrial networks, routing in a single network segment is not an issue at all. Fig. 2.6 shows the ISO/OSI model of Profibus DP as an example of a wireline industrial network. Absence of the Network Layer (Layer 3), which is responsible for traffic routing, is due to the fact that all of the nodes have direct access to a shared bus. If one of them talks, all of the others in the same segment can hear. Therefore, if the nodes have unique addresses in a segment, they could ping each other directly without the need of intermediate nodes. This holds true for all wireline single-segment industrial buses.

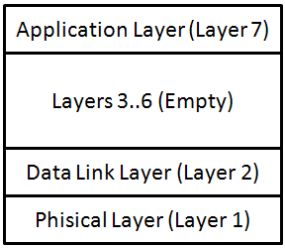

Figure 2.6: The ISO/OSI model of ProfiBus DP

As of wireless networks, if the coverage range of the transmitting node is adequate to reach the receiving node, relaying messages among different hops is not necessary, hence no routing protocol is needed. But an extensive coverage range has some drawbacks as well. It makes the system prone to more interference and collisions at the MAC layer because the wireless channel is broadcast in nature and the channel will be occupied in the vicinity of the transmitting node. The larger coverage range increases node density and heats up competition for gaining control of the channel. Moreover, it requires higher transmission power which is restricted by regulations. These two issues limit the coverage range of wireless nodes. In this project, we have assumed that routing is an indispensable part of wireless PnP control systems.

The problem of routing consists of two tasks which are: 1) characterizing nodes with some sort of information or properties that can be used by a routing algorithm, and 2) finding a routing algorithm. Reference [WMRD11] offers an up-to-date chronological survey on routing protocols for WSNs. In what follows, we have categorized those protocols in a different way to highlight different approaches in carrying out the above mentioned two tasks.

In terms of making a structure for the network, the following families of methods are identified in [WMRD11]:

1. No structure
2. Structuring based on geographical, relative or virtual coordinates
3. Structuring based on communication metrics
4. Application oriented structuring

Routing algorithms, on the other hand, can be generally categorized into three groups:

1. Routing based on a known sequence of addresses
2. Flooding based routing
3. Gradient based routing

A complete routing solution is a combination of a structuring method and a routing algorithm. The possible combinations are depicted in Fig. 2.7 and explained in details thereafter.

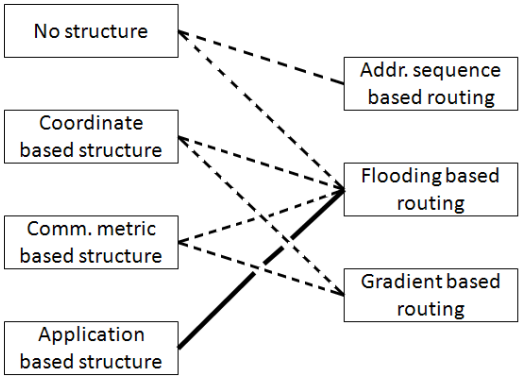

Figure 2.7: Possible combinations of structuring methods and routing algorithms to form a complete routing solution. The bold line shows the area of contribution of this thesis.

2.3.1 Routing based on a known sequence of addresses

This is actually the most obvious way to route a packet to its destination. It is originally devised for multi-segment wireline networks. The packet contains a sequence of node addresses and it is forwarded from each node to the next one whose address is known. The problem lies in how to devise a working sequence of addresses in the source node.

A popular approach is to flood *route request* packets through the network before or when a transmission is to happen. The sequence address will be decided based on the replies to route request messages. There are a number of well-known routing protocols which use this principle including Dynamic Source Routing (DSR) [JMB01], Ad-hoc On Demand Vector routing (AODV) [PBRD03], and Dynamic Mobile On-demand routing (DYMO) [CP10]. A simple and clear comparison between them is offered in [WMRD11].

2.3.2 Flooding-based routing

Chronologically, flooding is the first type of routing solutions that has appeared in wireless networks. Its name describes well how it works. When a node receives a packet, it checks whether it is the final recipient of the packet and if it has received the same packet before. In pure flooding, i.e. flooding with no structure, a negative answer to both of these questions results in retransmission of the packet. In this way, a packet is flooded throughout the entire network and will be received by every node in a connected segment.

Pure flooding has several advantages. First of all, it is a reliable routing method. If there is a good path from source to destination, the packet will eventually find its way and arrives at the final destination as soon as possible. Pure flooding is also the most robust routing method against topological modifications because there is no structure in it that needs to be updated after a topological change. But at the same time, it is painfully resource consuming. Too many healthy retransmissions are required before a packet is faded from the face of the network. This will not only consume much energy, which is valuable especially in battery-operated nodes, but also will increase conflicts and collision chance at the MAC layer.

Several remedies are proposed to reduce the number of unnecessary retransmissions in flooding-based routing. Some assume that a limited number of retransmissions is enough to reach the destination and do not propagate the packet any further. Such a number is derived either probabilistically [HEB08] or by considering the worst case [CJ03]. Others impose a structure to the network. In coordinate-based structuring, information either about the physical location or the virtual location of nodes in a system hierarchy is used to form clusters and flood packets only among cluster members [HCB02]. Communication metrics might also be involved in setting a structure for the network e.g. by omitting the nodes that have little residual energy or by blacklisting the paths which are not reliable enough [SR02]. The last method, i.e. application-based structuring, is the most interesting one for us in this thesis. The key in this method is to aggregate redundant data and transfer the minimal data packet that really affects the application. The available literature in this area proposes exploiting similarity of data content and basically dropping out redundant data [KHB02]. In the next chapter, we are going to offer a flooding-based routing solution which makes use of data content of packets to form application oriented clusters. More details are presented in the next chapter.

2.3.3 Gradient-based routing

Gradient based routing utilizes structure of the network to greedily relay packets towards one or a few number of data sink nodes. This approach implies assigning a scalar valued – per sink – to all nodes, representing their *distance* from that sink. Distance, could be any kind of routing metric [NTWe]. A route is then selected between the transmitter and the receiver nodes such that the shortest distance is taken between the nodes. The principle is to relay packets along the paths that provide the steepest gradient with respect to the network structure. It inherently assumes a dominant data traffic pattern, i.e. multi-point to point (MP2P) convergecast traffic. Most of the newer routing protocols in WSNs fall in this category which confirms the paradigm shift from adhoc peer-to-peer (P2P) to the structured MP2P convergecast traffic.

Most of the gradient routing methods are coordinate based. They require all of the nodes to know either their geographical, relative, or virtually defined locations by any of the following ways:

- In case of fixed locations, nodes could be pre-programmed with their coordinates in the network setup phase.

- Nodes could be equipped with the GPS navigation system to find their geographical location [IN99].

- Nodes could be in range of several anchor nodes to infer their relative location [BBG05] by localization techniques mainly *difference in time of arrival*.
- It is also possible to gradually find unaltering virtual locations of nodes by propagation of their *ranks* from the sink node toward their neighbors, neighbors of neighbors and so on [CWLY06].

The first and last positioning methods are only useful when nodes are not moving. To explain the last method, each node gets a scalar-valued discrete rank. The rank of the sink is zero and every other node chooses its own rank using the following equation.

$$R_i = min\{R_{i1}, R_{i2}, ..., R_{iK_i}\} + 1$$

where R stands for rank and i is the index of the node. R_{ij} is the rank of the j^{th} 1-hop neighbor of the i$^{\text{th}}$ node, and K_i is the number of 1-hop neighbors of the i$^{\text{th}}$ node. Using the above formula iteratively, results in convergence of nodes ranks to their minimal value [CWLY06].

After assigning location of nodes by any of the four bullet-pointed methods, the coordinates of the destination is stored in each packet before leaving the source node. Then it can be forwarded towards the destination either by a greedy approach or a combination of greedy and circumvolutory modes to escape network holes, which like local minima in optimization problems entrap the algorithm, deprive packet of being delivered in this context. Using combination of greedy and circumvolutory approaches guarantees delivery of packets to the final destination [FS06].

Other gradient-based routing methods utilize various communication metrics in order to define and assign nodes ranks. A comprehensive example is the under progress routing standard by Internet Engineering Task Force (IETF), namely Routing Over Low power and Lossy networks (ROLL) [NTWa]. IETF ROLL is the pioneer among routing standards since standardization activities used to be limited to lower layers of the ISO/OSI model including PHY and MAC. It is based on the ratified *tree* structure which may utilize various communication metrics in order to prioritize links or nodes in an accumulative or individual way [NTWc].

IETF ROLL is mainly designed for MP2P traffic. However, other traffic patterns of point to multi-point (P2MP) and P2P are also supported in the IETF ROLL. This means that P2P and P2MP are exceptions, only established after a node's request, and maintained temporarily [NTWd]. IETF ROLL is designed for monitoring and open loop control applications, as introduced in [NTWb], in which, one or a few equivalent sink nodes serve either as a data collecting point, a central controller, or an operator's station.

To select an appropriate starting point, the key question is whether MP2P traffic pattern serves closed loop control applications of P^{3}C case studies well enough. If the answer is affirmative, newer gradient based routing protocols could be utilized. Otherwise, we have to adopt one of the older routing solutions which is compatible with P2P traffic pattern. In this thesis, we have assumed that neither sensors nor actuators outnumber each other in comprising PnP control systems. This topological assumption makes gradient-based routing incompatible with the required wireless networks for PnP control systems. Therefore, we will focus on flooding-based routing in the next chapter.

3 | Methodology

This chapter summarizes the efforts that have been carried out and the results that have been achieved to answer the research questions that are already introduced in Chapter 1. It is divided into two subsequent sections, each of which dealing with one of the research questions.

3.1 Physical Communication

The first step to determine scope of the problem was to estimate the statistical distribution of distances among nodes of the network. Fig. 3.1 shows a typical urban area with district heating pipes. Each line represents a pair of pipes, i.e. both the forward and return pipes. Pumps are shown as small red dots on the pipes. Connecting pipes are buried about one meter below the surface along roads in sand filled channels. Pipes are made of ferromagnetic steel. They are all welded together and thermally insulated from the environment with polyester jackets filled with a thermal insulator [DHS].

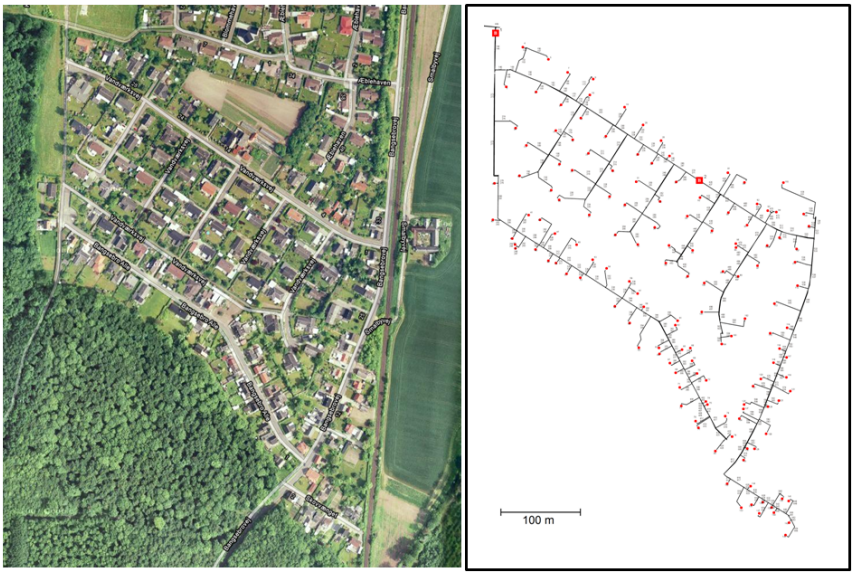

Figure 3.1: Satellite image of a sparsely populated urban area (Left) and its associated district heating piping layout (Right)

In a real scenario like the one depicted in Fig. 3.1, physical location of 180 pumps in a sample area was analyzed. Fig. 3.2 shows the cumulative distribution function (CDF) of distances for different number of nearest neighbors from one to five.

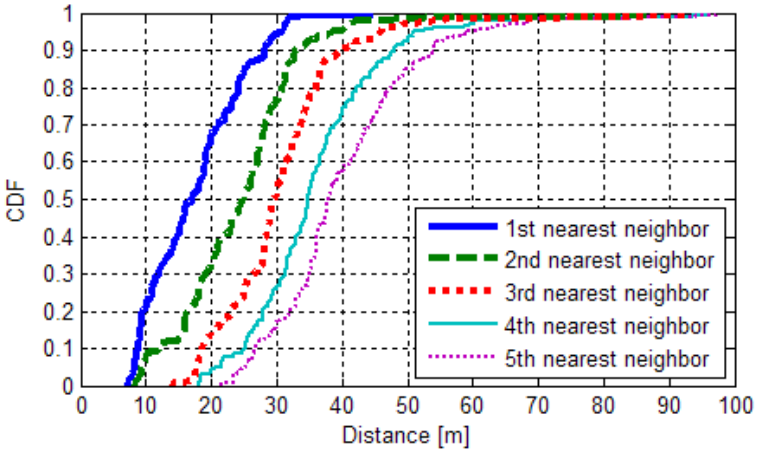

Figure 3.2: CDF of the first five nearest neighbor nodes

For instance, it shows that the distance for 50% of the pumps from their respective nearest neighbor is about 17 m. This increases to 25 m for the second nearest neighbor and so on. Another interpretation says that at 30 m distance from all of the pumps, there is at least one neighbor. It increases to 58 m for having at least two neighbors. Although having more neighbors at hand is beneficial in providing better connectivity, it will not guarantee a totally connected network. That is because even if all of the nodes have i nearest neighbors in reach, there still may exist isolated segments each one consisting of at least $i + 1$ nodes. In summary, a minimum required coverage range of 40-70 m could be a good practical starting point.

If the ubiquitous radio frequency electromagnetic wave technology was reliable enough, it would certainly be the best choice due to unprecedented availability and support of embedded radio chips. But reviewing the literature ended up in a disappointing outcome [AS06, ASV09]. Embedded radios are useful in wet soil medium only if the required range is less than 5 m. As a consequence, a survey on various potential technologies started.

Different methods were used to assess applicability of each of the studied physical communication technologies. The first impressions and availability of facilities were also influential on the extensiveness of investigations. For example, to see if *the electrical conductivity of pipes* can be exploited, only reading a technical report [EMS$^+$05] and comparing the conditions with the district heating case study seemed to be adequate, hence no further investigations were made in this case. On the other hand, the idea of using *acoustic waves* appeared to be quite promising in the beginning. Therefore, a more indepth literature survey was carried out both by considering pipes as the communication medium [LCL03, Kon05, Kok06, SNU06] and also in-pipe water as the medium [Sto08, WHD$^+$08]. However, the results were the same.

For some other technologies that seemed more promising and familiar as well, simula-

tions and experimental test were performed to assess their usefulness. Examples include, *steel-core magnetic induction*, and *pipelines as electromagnetic waveguides*. Simulation results for the former case were inspired by [SA09, BT80, Pop03, WCH88, Pop01]. As of the latter case, an experiment was designed and run.

Other technologies like *cell phone infrastructure* and *power line communications* are evolving very fast today. As a consequence, the current judgment that is made based on the latest available literature [Arr81, SZ84, Cla98, Ols02, GSW11] might not be valid in near future, especially in the case of power line communications which is a hot topic along with the emerging SmartGrid.

At the end, *air-core magnetic induction* seemed to satisfy all of the requirements. The method is not novel. It is already proposed [WL19, Bat20, Moo51, SWD01, Bun01, Ban04] and implemented for different applications [SGW99, Loc01, Loc04, RVD06, PCB07]. To examine its applicability, simulations were carried out to calculate the strength of propagated signal at relevant distances. Simulations were conducted by exploitation of analytical formulas for calculating magnetic field around coil antennas [Kun07]. The results followed by distinct optimization of transmitter and receiver antennas were promising with an ample margin. Successful simulation was a motive to conduct further investigations by designing and running experiments which turned out to follow simulation results with an acceptable margin.

3.1.1 Assessment of Potential Signal Propagation Techniques

Several signal propagation techniques were introduced in the last chapter. In this chapter, feasibility assessment results are reported.

As of exploiting RF EM waves in combined above-ground / underground scenario introduced in 2.1.1, empirically derived pathloss models in [KMH+08], which are valid for the operating frequencies of 2-6 GHz, were used estimate pathloss as a function of distance between transceivers for a set of frequencies. In our problem setting, pathloss (PL) can be estimated as follows:

$$PL[dB] = \quad 20\log_{10}(d[m]) + 46.4 + 20\log_{10}(\tfrac{f[GHz]}{5.0}) + 12n_w \tag{3.1}$$

where d is the distance between transceivers, f is the operating frequency, and n_w is the number of thick walls [KMH+08]. The results are exhibited in Fig. 3.3.

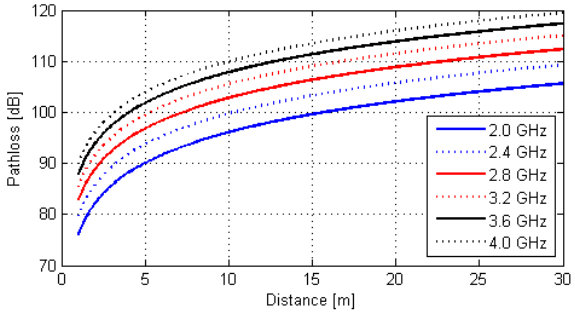

Figure 3.3: Pathloss vs. horizontal distance for several frequencies

Fig. 3.3 suggests a maximum transmission range of 15 m in ideal conditions, assuming a maximum tolerable signal loss of 100 dB, which is the dynamic range of current embedded radios. It is less than the minimum required range of 40 m, but it also shows that there could be a chance if lower frequencies are employed. However, (3.1) is not valid for f <2.0 GHz.

To see if lower frequencies can be useful, we ran a test with APC220 radio modules tuned at 434 MHz. The selected frequency is the lowest unlicensed frequency that could be supported by APC220 radio. The sensitivity of this radio is -112 dB at 9600 bps, and it uses GFSK modulation.

The experiment was a simple broadcast test. The transmitter was fixed in a basement, broadcasting a life signal once every second. The receiver, which was a mobile node, was listening to the life signals. We started with both nodes placed close to each other, then got out of the basement with the receiver node and moved towards the neighboring basement located in 10 m distance. It turned out that as soon as we were to go down in the second basement, the life signal was completely lost. This result is due to lack of the required reflecting surfaces. This approach was not pursued any further.

As of utilizing pipes as EM waveguides, an experiment was carried out to assess feasibility. The experiment setup is shown in Fig. 3.4. A 2.5" iron pipe conforming to DIN2440 standard was used in this test. It was integrated into the potable water distribution pipelines. As the transmitter and the receiver, two coil antennas were used each one made of 83 turns of single layered insulated copper wire conforming to AWG 30 standard. Lengths of the coils turned out to be 40 mm. Both antennas were shielded in cardboard boxes which were totally covered by a thick enough aluminum foil. Therefore, the signal could only transmit through the pipe.

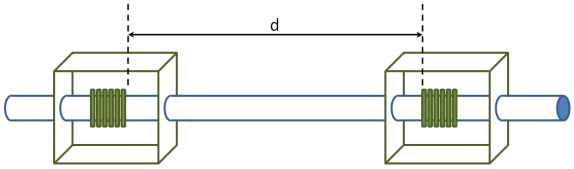

Figure 3.4: Metallic waveguide experiment setup

The coils were used with a *Signal Generator* and *Spectrum Analyzer* without impedance matching circuits. Fig. 3.5 shows signal loss as a function of distance (d) between the coils for operating frequency 400 MHz. Impedance mismatch of the two circuits imposes a large constant signal loss, but we were generally interested in the variations of signal loss versus d. Therefore, it was possible to ignore the matching problem assuming that the matching parameters were not affected considerably while d varied.

The oscillatory behavior of signal loss admits existence of the waveguide effect. The observed wavelength is approximately 0.5-0.6 m which is in accordance with the calculated wavelength for 400 MHz in water medium, i.e. 0.56 m, provided that:

$$v = \frac{c}{n} \qquad (3.2)$$
$$\lambda = \frac{v}{f}$$

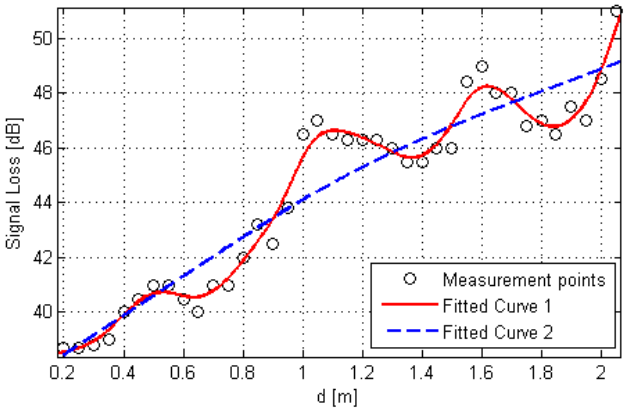

Figure 3.5: Waveguide effect on signal loss at 400 MHz

where c (2.9979×10^8 m/s) is the speed of light in vacuum, n (1.333) is the *index of refraction* of water, v is the speed of light in water, λ is the wavelength, and f is the operating frequency [NB85].

Measured data are smoothed and plotted in Fig. 3.5 in two steps. The first fit explicitly shows the oscillatory behavior, while the second fit shows the general trend and anticipates, via extrapolation, 100 dB of signal loss at approximately 20 m after impedance matching. Unfortunately, we did not have access to longer straight water filled pipes to verify this result, but it is not satisfying for our DHS problem yet, considering the required range of 40 m and the presence of bends and branches in pipes in practice, which severely affect the waveguide properties.

As of exploiting acoustic waves and the piping infrastructure as signal carriers in a practical economic setup, we performed a simple test with fairly cheap piezoelectric buzzers used as contact microphone and speaker. The buzzers were fixed with epoxy glue on an iron air-filled open-ended pipe. Single carrier transmission test at 4.1 kHz, which was the piezoelectric resonance frequency, showed that a maximum 10 v input voltage at the transmitting buzzer produces less than 0.1 mv at the receiving buzzer at a distance of 0.5 m. Although the sonic signal could be heard clearly, it was not detectable by the oscilloscope for larger distances. In summary, it was understood that a microphone with a high sensitivity, should be used and it costs too much even in bulk numbers.

Regarding acoustic solutions, it should be noted that the communication framework that we are interested to build is going to host a control system for regulating pressure and compensating pressure disturbances in a DHS. This means that the transmitted data have the highest importance when pressure irregularities happen, but these irregularities also induce vibrational noise which degrades reliability of the communication link. Therefore, the channel is least reliable when it is most needed. Marginal reported experimental results for short pipes, the high price of the required hardware, i.e. comparable to the price of the pump itself, and the effect of the process disturbance on the communication channel makes the current acoustic technologies unreliable for in-pipe communications.

The last signal propagation technique that we studied beyond the available literature was Magnetic Induction. Both through the pipes and through the soil options were considered. In the former case, self and mutual inductances were used to find the ratio between the input complex power to the transmitter coil (P_t) and the output complex power from

an impedance load (P_r) in order to evaluate MI pathloss. Using the equivalent electric circuit of coupled inductors, and matching the load impedance to receive the highest power, the exact expression for power transmission ratio is derived in [SA09] as:

$$\frac{P_r}{P_t} = \frac{\left[j\omega M/(R_t + j\omega L_t)\right]^2}{4R_r + 4R_t\omega^2 M^2/(R_t^2 + \omega^2 L_t^2)} \cdot \left(R_t + j\omega L_t + \frac{\omega^2 M^2}{2R_r + \omega^2 M^2/(R_t - j\omega L_t)}\right) \tag{3.3}$$

where ω is the angular velocity of the signal, and R_t and R_r are the electrical resistance of coils. Evaluating (3.3) was the final step to assess the feasibility of through-the-pipes MI communication. For this purpose, the CST EM Studio®magneto-static solver was used for finding self and mutual inductances. Some parameters of the pipe were: 1) outer diameter 38 mm, and 2) thickness 3.65 mm, conforming to DIN 2440 standard. Pipes were filled with water.

While self inductance was almost independent of inter-coils distance, mutual inductance varies according to plotted simulation results in Fig. 3.6. Consequently, while [ASV09] and [SA09] predict a range no better than 3 m for 100 dB admissible pathloss, Fig. 3.6 promises enhanced ranges by showing much higher mutual inductance compared to air-core coils.

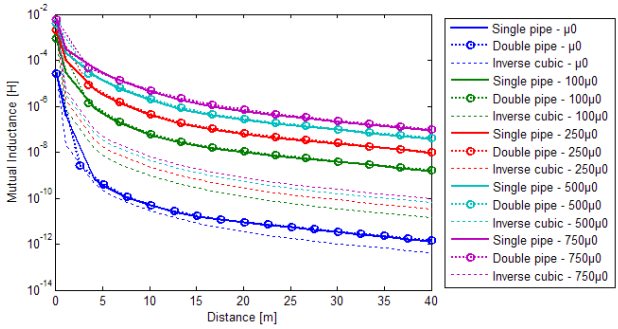

Figure 3.6: Comparing the diminution of mutual inductance M with $1/r^3$

The next step is to evaluate (3.3) and calculate pathloss between the transmitter and receiver coils. Fig. 3.7 shows the pathloss, for operating frequency 3000 Hz. It is assumed that the wire unit length resistance is 0.34 ω/m, conforming to AWG 30 standard.

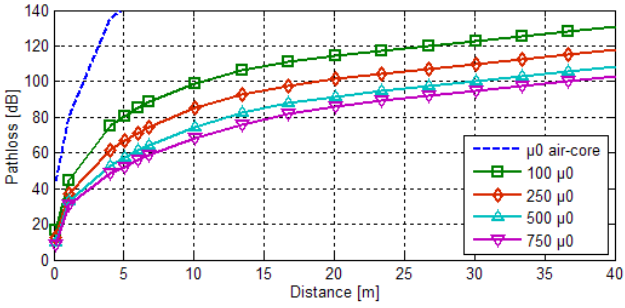

Figure 3.7: Pathloss versus distance between coils at 3000 Hz

The dotted line represents the air-core case, where having a viable communication link of approximately 2.5 m is expected. The coverage range is considerably increased by using a ferromagnetic core. Magnetic permeability of ferromagnetic steel pipes varies in a wide range up to several hundred μ_0 depending on their metallurgic composition. This gives a maximum coverage range from approximately 20 to 40 m for $250\mu_0$ to $750\mu_0$, provided that a maximum 100 dB of pathloss is tolerable.

Although the simulation results are marginally acceptable, practical considerations like bends in pipes, branches and fittings all produce detrimental effects. Moreover, eddy current losses, B-H nonlinearity and B-H frequency dependent hysteresis losses contribute to degrade signal strength even at frequencies as low as 3 kHz.

In the other approach, where through-the-soil MI was employed, non-radiating fields as described in **2.1.6** were utilized to make communication links.

Dielectric properties of lossy materials like wet soil encouraged us to use low frequency non-radiating fields. When materials are influenced by alternating $\mathbf{E}$ and $\mathbf{B}$ fields, their electric susceptibility (χ_e) can be stated as a complex function of frequency:

$$\chi_e(\omega) = \chi'_e(\omega) - j\chi''_e(\omega) \tag{3.4}$$

The general behavior of real part $\chi'_e(\omega)$ and imaginary part $\chi''_e(\omega)$ for most of the materials is sketched in Fig. 3.8 [Bal89]. $\chi'_e(\omega)$ increases most of the times with respect to ω, except for short intervals where it drastically decreases. These intervals coincide with local maxima in $\chi''_e(\omega)$. $\chi'_e(\omega)$ determines velocity of $\mathbf{E}$ and $\mathbf{B}$ fields in a matter, while $\chi''_e(\omega)$ accounts for attenuation due to molecular resonance phenomena.

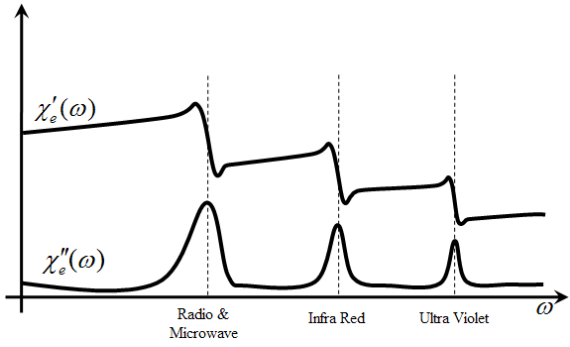

Figure 3.8: Electrical susceptibility of matters versus frequency

It provides an extra *alternating field conductivity* (σ_a) besides the common *static field conductivity* (σ_s) at resonance frequencies. The overall conductivity (σ_e) can be stated as [Bal89]:

$$\sigma_e = \sigma_s + \sigma_a = \sigma_s + \omega(\epsilon_0 \chi''_e(\omega)) \tag{3.5}$$

Equation (3.5) and Fig. 3.8 imply that, $\mathbf{E}$ and $\mathbf{B}$ fields should have frequencies lower than the common radio and microwave bands, i.e. the first peak in $\chi''_e(\omega)$ graph, in order to decrease overall conductivity, hence attenuation. Furthermore, in non-radiating fields, there holds no relation such as $|\mathbf{E}|/|\mathbf{B}| = v$. Therefore, $\mathbf{B}$ can be substantially large. It also means that the energy does not have to be divided equally between the $\mathbf{E}$ and $\mathbf{B}$ fields, as it does in radiating fields.

3.1.2 Simulation Results on MI Coil Antenna Design

Our initial studies showed that there could be a chance for MI to be viably employed as the signal propagation technique in our underground DHS network. However, some considerations in the antenna design procedure should be taken into account. Fig. 3.9 shows the basic elements of a MI communication system. Coil antennas are used at the transmitter (Tx) and receiver (Rx) sides. Two coil transceivers operating at low frequencies can be

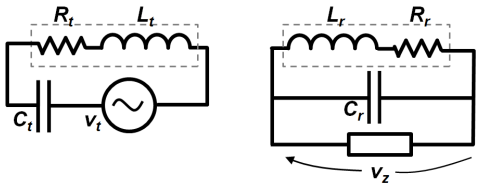

Figure 3.9: MI transmitter and receiver circuits

viewed as a core-less transformer with two separate coils and inefficient coupling. Both coils can be fairly modeled as an ideal inductor in series with a resistor which is equal to resistance of the coil wire. We assume that the operating frequency is low enough, hence ignore the self capacitance across the coil in its lumped model.

At Tx side, the objective is to produce the strongest possible $\mathbf{B}$ field. This will be achieved when electrical current in the coil (I_t) is maximized. Therefore, a capacitor C_t should be connected to the coil in series and bring it to resonance such that $I_t = v_t/R_t$. In order to increase I_t further, R_t should be reduced. It can be achieved by breaking the coil into several parallel coils, connecting some/all of the individual wire turns in parallel [Loc01]. If the total DC resistance is R and the wire turns are to be grouped in N parallel coils, each with DC resistance of R/N, the overall resistance (R_t) will be equal to R/N^2.

The Rx coil, on the other hand, should be sensitive to the slightest change of the passing magnetic flux and produce the highest back-emf. The induced RMS emf v_r can be calculated in a multi-turn coil by:

$$v_r = \omega . \sum_{i=1}^{N_r} \left[\int_{A_i} \bar{\mathbf{B}} . \hat{\mathbf{n}} d(s) \right] \tag{3.6}$$

where ω is the angular frequency of I_t, A_i the area of individual turns of the Rx coil, N_r the number of turns of the Rx coil, and $\bar{\mathbf{B}}$ is the RMS magnetic field passing through the it. If $\bar{\mathbf{B}}$ has a constant amplitude at all the points on the Rx coil surface and is parallel to $\hat{\mathbf{n}}$, which has unit length and is perpendicular to the coil surface, (3.6) will be simplified as follows:

$$v_r = \omega . |\bar{\mathbf{B}}| . \sum_{i=1}^{N_r} A_i \tag{3.7}$$

which shows that $\sum_{i=1}^{N_r} A_i$ should be maximized in order to get the highest possible back-emf. Furthermore, if a capacitor C_r is connected to both ends of the coil such that

it brings the coil into resonance (see Fig. 3.9), the followings hold:

$$\begin{aligned} C_r &= \frac{1}{\omega^2 L_r} \\ v_z &\approx \frac{1}{jC_r\omega} I_r = -j\omega L_r (\frac{v_r}{R_r}) \\ &= -jQv_r \end{aligned} \tag{3.8}$$

provided that the input impedance of the receiver amplifier/filter is very large and the *quality factor* (Q) of the coil is defined as $Q = \omega L_r / R_r$. Therefore, if the Rx coil has a high Q, the passive band-pass receiver circuit acts as a pre-amplifier.

Combining (3.7) and (3.8), the design objective in Rx antenna is to maximize the following expression:

$$\frac{L_r}{R_r} . \sum_{i=1}^{N_r} A_i \tag{3.9}$$

Given a specific wire gauge R, and length l, an interesting problem is how to wind the coils such that the design criteria in Tx and Rx antennas are met. Assumptions on wire gauge and length are reasonable since they define cost and weight of the coils. They hold throughout this section.

We start by introducing the analytical expression for the $\mathbf{B}$ field in all points of the space, created by a single turn current loop as shown in Fig. 3.10. The formula is derived by calculating the magnetic vector potential first, and taking the curl of it to find $\mathbf{B}$ [Kun07].

$$\begin{aligned} \mathbf{B}(r,z) = \frac{\mu I k}{4\pi\sqrt{ar^3}} \Bigg[-(z-h)\left(K(k) - \frac{2-k^2}{2(1-k^2)}E(k)\right)\hat{\mathbf{r}} \\ + r\left(K(k) + \frac{k^2(r+a)-2r}{2r(1-k^2)}E(k)\right)\hat{\mathbf{z}} \Bigg] \end{aligned} \tag{3.10}$$

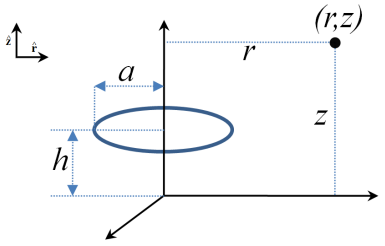

Figure 3.10: A current loop

$\mathbf{B}$ is expressed in terms of complete elliptic integral functions of first and second kind, i.e. $K(k)$ and $E(k)$, where their argument k is defined as:

$$k = \sqrt{\frac{4ar}{(r+a)^2 + (z-h)^2}}$$

All other parameters of (3.10) are expressed in Fig. 3.10. If $\mathbf{B}$ is only sought on z-axis, (3.10) will be significantly simplified. In that case, it is more convenient to directly use Biot-Savart law to calculate on-axis $\mathbf{B}$ field.

3.1.2.1 Propagation Pattern of Coil Antenna

The graphical representation of $|\mathbf{B}|$ for a multi-turn coil is given in Fig. 3.11 as a contour plot. The coil is not shown in this plot, but it is laid in $z = 0$ plane and centered at the origin, with $a = 0.1$ m. The field is calculated at each point in space by summing up (3.10) for individual wire turns, considering the exact radius and location for each turn of wire.

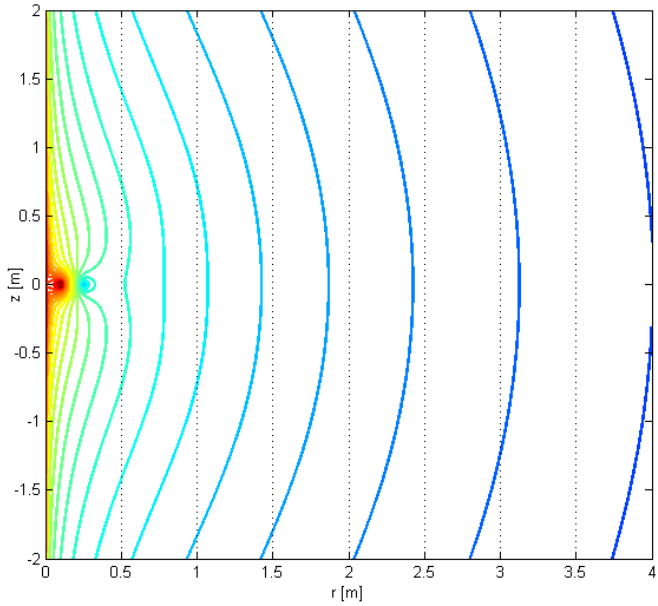

Figure 3.11: Contour plots of $|\mathbf{B}|$ in the vicinity of a transmitting coil antenna

Since the contours in Fig. 3.11 do not follow an exact spherical pattern, the antenna pattern is not omni-directional and relative placement of the Tx and Rx antennas should be studied. It is clear that $|\mathbf{B}|$ is larger around the z-axis. This behavior is shown quantitatively in Fig. 3.12, in which $|\mathbf{B}|$ is drawn at three different fixed distances of 1 m, 10 m, and 100 m from the Tx antenna. Each curve starts from vicinity of the z-axis, associated to $\theta = 0.1°$ to $z = 0$ plane, associated to $\theta = 90°$. The results are shown in log-log scale in sake of highlighting the differences. Moreover, the curves are shifted vertically such that they all coincide at $\theta = 90°$. Fig. 3.12 shows that $|\mathbf{B}|$ can vary up to approximately 20 dB at the Rx antenna based on positioning of the Tx antenna such that $\theta \in [1°, 90°]$. It also shows that $|\mathbf{B}|$ is more sensitive when θ is small.

Note that Fig. 3.12 does not say anything about the angles between Tx and Rx coils' axes, which is important in calculating the magnetic flux passing through the Rx coil.

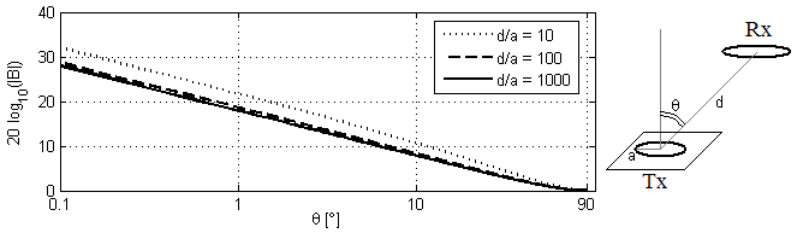

Figure 3.12: Normalized $|\mathbf{B}|$ vs. Rx coil position with respect to Tx coil axis

3.1.2.2 Cross section shapes

Given the mean radius of the coil a, and hence the number of its turns N, ($N = l/(2\pi a)$), we have compared rectangular cross section shapes with different side proportions to see its effect on the Tx and Rx antennas. These are ranging from a single layered coil in which $b = 2R$ and $c = Rl/(\pi a)$ to a coil with the maximum number of winding layers which is made when $b \leq 2a$ and $c \geq R^2 l/(\pi a^2)$. The latter holds, provided that each wire occupies a square surface of area $4R^2$ at the cross section of the coil. The most compact wiring, results in hexagonal surface assignment to each wire turn and $c \geq \frac{\sqrt{3}R^2 l}{2\pi a^2}$. All of the above inequalities convert to equalities if the wire is very long such that no hole remains in the center of the multi-turn coil.

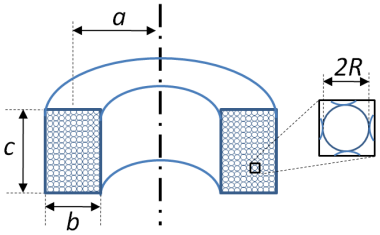

Figure 3.13: Dimensions of a multilayer coil with rectangular cross section

At Tx coil, the difference between $|\mathbf{B}|$ of a single layer and a square coil at $z = 0$ plane is in order of 0.01% of the $|\mathbf{B}|$ created by either of them. In all other points in the space, this trifling difference is even more negligible. Therefore, the cross section shape of the Tx antenna is not tangibly influential on $|\mathbf{B}|$.

At the Rx coil, on the other hand, the coil's cross section shape is quite effective mostly due to its effect on the self inductance of the coil (L_r). This is explicitly presented in (3.9). Since the length and the mean radius of the coil are predefined, $\sum_{i=1}^{N_r} A_i$ is constant and L_r must be maximized.

Maximization of self inductance of a coil was first posed by Maxwell in [Max73]. He has shown that the *geometric mean distance* between the conductors of an inductor should be minimized in order to get the maximum inductance. This is achieved when the cross section is circular. For rectangular shapes which are easier to manufacture, the best shape is a square, i.e. $b = c$ (See Fig. 3.13). This result is verified by simulations in Fig. 3.14.

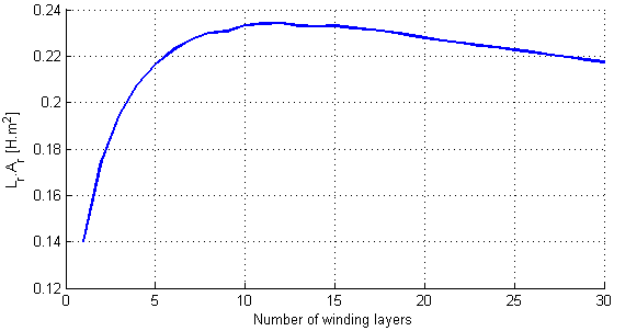

Figure 3.14: emf variations in the Rx coil as a function of cross section shape

The coil in this simulation is made up with 150 m of wire with gauge $R = 0.3$ mm. The cross section is rectangular and the mean radius is $a = 0.20$ m. Consequently, the number of turns N_r turns out to be 119 which yields 11 winding layers for a square cross section. Fig. 3.14 confirms this fact by showing a peak at N_r =11. It also shows that a single layer Rx coil has only 60% performance of the square shaped Rx coil.

In the above simulation, L_r is calculated by injecting a current I into the receiver coil and computing the induced magnetic flux Φ_i in each turn due to the current in all of the turns. Then, $L_i.I = \Phi_i$ gives L_i and $L_r = \sum_{i=1}^{N_r} L_i$. Note that, in calculating the total self inductance of the Rx coil, we only added up L_i of individual turns which convey the self inductance of that turn plus its mutual inductance to all other turns.

3.1.2.3 Mean radius vs. Number of turns

Since the total wire length ($l = 2\pi aN$) is predefined, aN is constant and we should make a trade-off between a and N. At the Tx antenna, Fig. 3.15 exhibits the variations in $|\mathbf{B}|$ at $z = 0$ as r increases from 1 m up to 50 m, for different values of a from 0.03 m to 0.25 m.The lowest value is associated with the physical limitation of $a > b/2$ (see Fig. 3.13 for clarification).

Fig. 3.15 suggests that a should be as small as possible to have the maximum $|\mathbf{B}|$ at large distances. The same results were achieved when $z \neq 0$ and also on any straight line with slope α such that $z = \alpha.r$. No matter how large a finite α is, the above hypothesis holds true and the smallest mean radius gives the maximum $|\mathbf{B}|$ at distances of interest. The only exception is $\alpha = \infty$, i.e. $|\mathbf{B}|$ on the z axis in which the converse is true, but it does not have any practical significance.

For the Rx coil, expression (3.9) is depicted in Fig. 3.16 as a function of a. For all values of a, a square cross section is considered.

Fig. 3.16 admits that for the Rx coil, larger a produces higher voltages, no matter if the number of turns or L_r is decreased significantly. In practice, a is usually specified by the space limitations. Moreover, Q of the Rx coil should be kept greater than 1, not to attenuate the signal. The presence of C_r capacitor is always necessary to adjust the phase of the received signal, such that it can be fed into amplifiers with a resistive input impedance.

In summary, the design criteria for coil antennas are as follows:

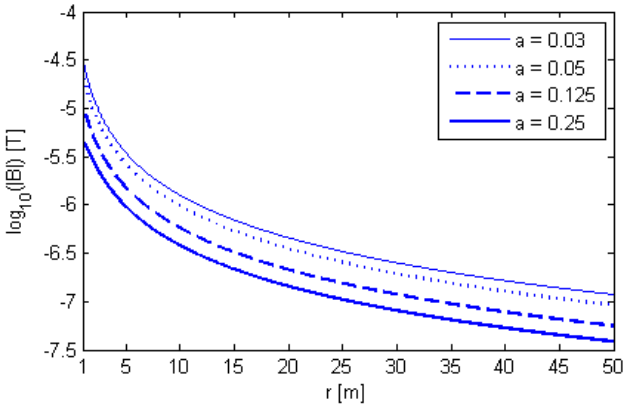

Figure 3.15: Variations in $|\mathbf{B}|$ as r increases

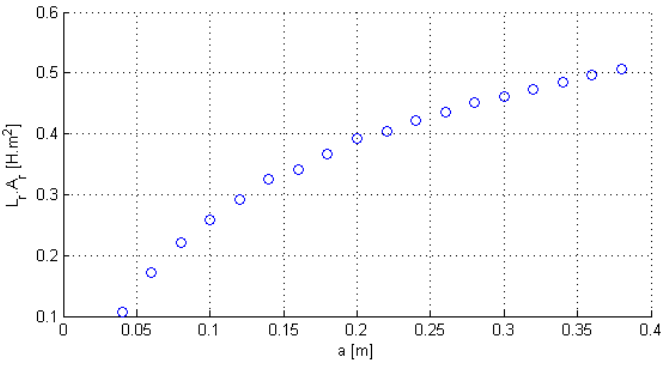

Figure 3.16: Variation in v_z at the receiver coil as a increases

- The Tx coil should have the smallest possible a while the Rx coil should have the largest possible a.
- The cross section of the Rx coil should be circular or square (when a circular shape is difficult to manufacture), while the cross section shape of the Tx coil might be freely chosen based on practical winding issues.

  Combining the above two statements suggests that the Tx coil should be single layered with a very large length (c) and $a = b/2$ which is impractical. Therefore, in practice, the maximum permissible length of the Tx antenna (c_{Max}) limits the size of the coil. As for the Rx coil, the maximum permissible radius (a_{Max}) specifies its size.
- The Rx coil center ought to be located on the Tx coil axis. Otherwise the attenuation in $|\mathbf{B}|$ could be approximated by the plot in Fig. 8.11.
- When the coils and their locations are known, the receiver should face the right direction such that $\mathbf{B}$ is perpendicular to the Rx coil surface. In worst case, when $\mathbf{B}$ and the coil's surface are parallel, no signal will be received.

3.1.3 MI Signal Propagation Experimental Results

Having found the best conditions for Tx and Rx antennas, we investigated feasibility of communications in our district heating system. Feasibility study is done by planning a specific simulation scenario and comparing its results with actual measurements from an experiment with the same specifications.

Three coils are used in our simulations and experiments with the following specifications listed in Tab. 3.1. Function of each coil is assigned with respect to the findings in the previous section.

Table 3.1: Coils specifications

Coil#	Coil 1	Coil 2	Coil 3
Wire length	148 m	359 m	142 m
Wire gauge	0.3 mm	0.35 mm	0.3 mm
Wire diameter incl. insulator	0.65 mm	0.80 mm	0.65 mm
Wire specific DC resistance	0.06 Ω/m	0.045 Ω/m	0.06 Ω/m
Wire measured DC resistance	8.90 Ω	16.19 Ω	8.90 Ω
Coil mean radius (a)	2.8 cm	30 cm	11.3 cm
Coil cross section shape	rect.	circ.	circ.
Coil width (b)	≈3.3 mm	≈12.5 mm	≈10.5 mm
Coil length (c)	111 mm	≈12.5 mm	≈10.5 mm
No. of wire turns	841	190	200
No. of winding layers	5	N/A	N/A
No. of wire turns per layer	169	N/A	N/A
Function	Tx	Rx	Tx/Rx

Remark 1: It is assumed that $\mathbf{B}$, unlike $\mathbf{E}$ has a similar attenuation in air and in wet soil due to similar magnetic permeability in these environments. This is explained by the practical separation of $\mathbf{B}$ and $\mathbf{E}$ fields at non-radiating fields domain. Therefore, we have carried out all of the experiments in office environment, not in wet soil.

Remark 2: Operating frequency is chosen to be 5055±1 Hz for all three coils. Variable capacitors are connected to each of the coils and adjusted accurately to attain 5055 Hz resonance frequency. However, since the experiments are carried out in office environment, it is observed that the self inductance of the coils, and hence their resonance frequency is dependent on surrounding ferromagnetic objects, e.g. office heating radiators. Therefore, re-adjustment of the capacitors is needed during experiments.

Remark 3: During experiments, an audio amplifier is used in connection with a digital signal generator at the Tx side to be able to provide the required electrical current. At the Rx side, an oscilloscope receives signals without any additional filter or amplifier.

3.1.3.1 Measured vs. calculated self inductance

The purpose of the first test was to verify if (3.10) in combination with the procedure introduced in **3.1.2.2** can be used to accurately calculate self inductance of a coil. Since only inductance of Rx coils are of interest in our simulations in the previous section, coil 2 and coil 3 are considered here. Calculated values are shown in Tab. 2 against actual values which are measured by two different methods. The first method involved direct

use of a RLC-meter to measure L_r. The second method employed $L = 1/(4\pi^2 f^2 C)$ to find L based on using a known capacitor (C) in a LC circuit with our coil and measuring its resonance frequency by oscilloscope. The capacitor was connected in series with the inductor in this test for more accurate measurements in practice.

Self inductance is of interest for the Rx antenna. Thus, the second and the third coils with larger mean radii are considered in this experiment. We have calculated their self inductance by the procedure introduced in **3.1.2.2**. The results are shown in Tab. 3.2 against the actual values which are measured with two different methods. The first method involves direct use of a RLC-meter to measure L_r, while the second one enjoys a higher accuracy by measuring the capacitor at a specific resonance frequency and employing $L = 1/(4\pi^2 f^2 C)$. The capacitor should be connected in series with inductor in this test. The results show 1% difference between the calculated and measured values.

Table 3.2: Measured versus calculated self inductance

Coil#	Calc. L	Direct Meas. L	Indirect Meas. L
2	55.11 mH	54.5 mH	54.44 mH
3	18.37 mH	18.17 mH	18.12 mH

3.1.3.2 Surrounding Ferromagnetic Objects

In this test, it was initially intended to verify that the received voltage across the Rx antenna (v_z) is a linear function of the electrical current in the Tx antenna (I_t) in practice, as it is evident from the Biot-Savart law. However, our observations suggest a different title for this experiment.

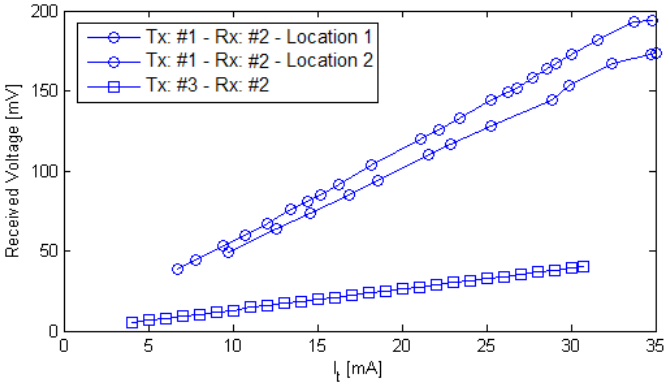

Figure 3.17: Received voltage as a function of Tx current

The test was planned for two combinations where either coil 1 or 3 are used as the Tx and coil 2 as the Rx antennas. Fig. 3.17 shows linear dependency between v_z and I_t for both cases. But the results were not easily reproducible. They varied if nearby ferromagnetic objects were replaced, e.g. office chairs with iron parts. To visualize this effect, two graphs for coil 1 as the Tx antenna are plotted. Placement of the surrounding objects is quite different in one case from the other. All other parameters are unchanged. It can be seen that the slope of the line is changed. This experiment raises some concerns about the effect of steel pipes, in district heating systems, on the $|\mathbf{B}|$ field and self inductance of

coils. More experimental studies or real scale implementations are needed to see whether this effect acts in favor or against our system in overall. The bottom line is that various physical piping layouts will impose an uncertainty in quality of the received signal.

3.1.3.3 Measured vs. calculated back-emf

This is the main experiment, in which we compared simulated and measured values of v_z. Based on (3.8) and backed by the promising results of Experiment 1, this test actually compares simulated and measured values of $|\mathbf{B}|$ at Rx side.

For coil 3 as the Rx antenna, (3.9) turned out to be 1.78572×10^{-2} H.m^2/Ω. This value multiplied by ω^2, where the operating frequency is 5055 Hz, gives 1.80142×10^{7} V/T which is voltage gain of the Rx coil. It is equal to the RMS of v_z if assuming: 1) a uniform $|\mathbf{B}|$ on surface of the Rx antenna, and 2) the ideal direction of the Rx antenna such that its surface is perpendicular to the $\mathbf{B}$ field.

Simulations and experiments are performed on $z = r$ cone. I_t is chosen such that peak-to-peak voltage across C_t does not exceed 40 V, i.e. the maximum permissible voltage for our capacitors. I_t turned out to be 31.0 mA and 28.8 mA RMS for coil 1 and coil 3, respectively.

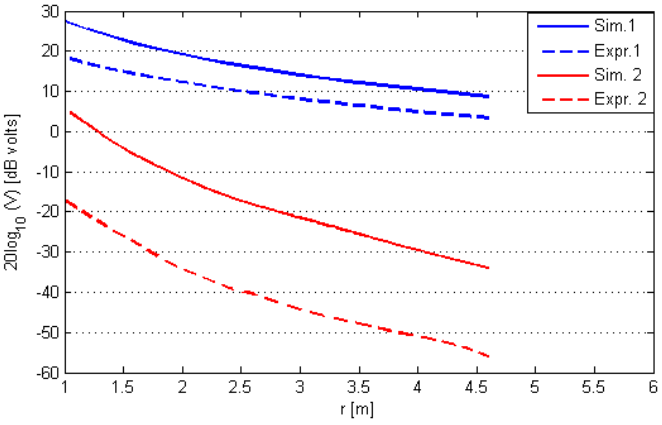

Figure 3.18: Calculated versus measured received voltage v_z for two Tx coils

Calculated and actual voltages are shown in Fig. 3.18. It is obvious that coil 1, as the Tx antenna, is superior to Coil 3 both in simulations and in practice. If we consider a detection threshold of 1 mV, i.e. -60 dBm, the configuration with coils 3 and 2 is useful for no longer than 5 m links. However, the other configuration with coils 1 and 2 looks quite promising at this range. Even better results can be achieved by following coil antenna design criteria of the previous section, more precisely.

Unfortunately, the phenomenon that we observed in Experiment 2, affected this experiment as well. When distance increased, the results became less reproducible. That is why we have not plotted the graph for longer ranges. In summary, the results are sensitive to presence of ferromagnetic objects between transceivers. Effects are not always destructive, and more advanced experiments are needed in order to commercialize this solution. Concluding remarks on MI signal propagation are in the next chapter.

3.2 Networks and Protocols

This section tries to answer the second research question of the project posed in **1.5**. The objective is to select or devise suitable MAC and routing protocols based on the reviewed literature in **2.2** and **2.3**. It is also an advantage if the selected protocols for MAC and NTW layers create a clear interface with the control application on top of the ISO/OSI stack. To this end, a control oriented routing metric is used in design of the routing protocol. Therefore, the control application has a direct influence on the networking topology.

In what follows, we studied the predominant traffic pattern and chose the networking topology. Assumptions on the network topology led us to pick flooding as the routing scheme. Details of the routing algorithm and introduction of the new routing metric follow afterwards.

3.2.1 Dominant Traffic Pattern

A WPnP control system consists of three kinds of nodes: actuators, controllers, and sensors. All of them should have the capability to act both as a data *sink* and as a data *source*, described as follows.

- A Sensor is regularly a data source to send sensory data towards relevant controllers, typically once per control *cycle time* interval, equivalent to the control loop sampling time, e.g. 100 ms.
- A sensor sporadically acts as a data sink to receive configuration data from its associated controllers.
- An actuator is regularly a data sink which receives commands from the associated controller at each control cycle time and implements them.
- An actuator sporadically acts as a data source to report failures.
- Controllers should send and receive data in each control cycle time interval. They gather data from sensors at the beginning of a typical cycle time interval, and send commands to the actuators at the end of the interval. This is exactly what happens in a Programmable Logic Controller (PLC), supplied either with individual input/output cables or aggregated bus communication modules.

Unlike newer routing protocols, which assume a Multi-point to Point (MP2P) traffic with controllers as the sink nodes [NTWc], the above list suggests a P2P traffic pattern. It is due to the assumption that neither sensors nor actuators outnumber each other significantly in WPnP control systems. Therefore, WPnP control systems are incompatible with MP2P architecture and solutions, e.g. gradient-based routing. Note that, this would not be the case if only monitoring and open loop control were of interest as in the IETF RPL[NTWc].

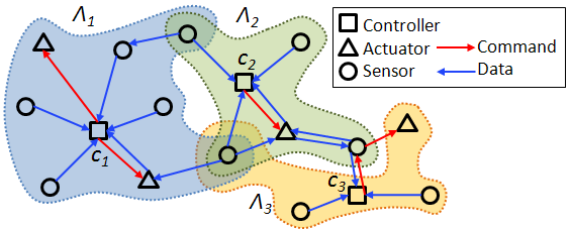

Figure 3.19: Logical network topology of a simple decentralized wireless networked control system

3.2.2 Network Topology

Fig. 3.19 shows the logical network topology of a WPnP control system with multiple controller nodes. Arrows represent data direction in the regular functioning mode, but information flow in the reverse direction is also required sporadically as explained in the previous section. Fig. 3.19 illustrates the key assumptions that are considered in topology design, described in the following:

1. There could be a large number of controllers ($C_i, i = 1, ..., n$).

2. Each controller and its associated sensors and actuators form a set, called a *cluster* henceforth. Clusters are identified by their unique tag ($\Lambda_i, i = 1...n$).

3. There are as many clusters as controller nodes which are called *cluster heads*.

4. Each data packet in the network layer contains a cluster *association* field. In general, a packet might be tied to one or more clusters. It is also possible that a packet is not associated with any cluster.

5. A sensor might be a *member* of multiple clusters, meaning that its generated data could be associated to more than one cluster head. In other words, a packet that is generated at a sensor node, might be reported to more than one controller.

6. An actuator could be a member of at most one cluster, meaning that it may not receive commands from more than one controller.

7. Data packets to/from members of a cluster should be sent from/to the cluster head. In other words, a controller node is either the source or the final destination in every transmission path.

8. Relaying data packets associated to the cluster (Λ_j) could be done via all of the nodes in the entire network, irrespective of their membership status in Λ_j, provided that the relayed packet has enough remaining *relay credit*.

Relay credit can be defined in terms of any scalar node or link routing metric, e.g. *hop count* as a node metric or *expected transmission count (ETX)* as an accumulative link metric [NTWe].

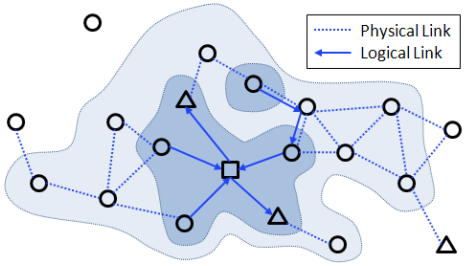

Figure 3.20: Preserving connectivity of cluster members by a surrounding cloud whose thickness is adjusted by *relay credit*, here defined as hop-count ≤ 2

The last assumption clarifies that cluster membership is not essential when relaying a packet. The purpose of assigning relay credit to a packet is to preserve connectivity in a cluster while constraining the number of retransmissions that might occur to a packet among non-member nodes. Each data packet is given an initial relay credit besides cluster membership tags, when generated. While it is roaming inside its own cluster, it does not spend any relay credit. However, when it is being relayed among non-member nodes, the initial relay credit is decreased at each non-member node until the remaining credit is not enough for more retransmissions among non-member nodes. Fig. 3.20 illustrates an example.

In Fig. 3.20, the dark blue area includes members of a cluster Λ, and the surrounding light blue area shows the maximum penetration depth of packets from Λ if hop-count ≤ 2 is considered as the relay credit criterion. If the light blue area is a connected region despite the fact that there is not a single connected dark blue region, connectivity still holds. Relay credit helps preserving connectivity of Λ members too some extent, when there is no direct physical link between them. This is shown in Fig. 3.19 too, where a sensor node which is a member of Λ_3 connects to its own cluster via an actuator and a sensor of Λ_2.

3.2.3 A Novel Application-based Routing Metric

Inapplicability of gradient-based routing algorithms, large routing overhead of unconstrained flooding-based algorithms, and the necessity of providing a clear link between the networking layer and the application layer, encouraged us to develop a way of clustering nodes that originated from requirements of the application layer and can efficiently constrain the domain of packet propagation in a clustered flooding-based routing algorithm. To this end, we are going to make use of data content of packets in a clustering algorithm.

Current solutions for content-based clustering are either based on tailoring redundant data or aggregating correlated or similar data [KHB02]. Our solution utilizes control oriented metrics to form clusters, instead of typically used communication based metrics. It works by classifying the nodes into different clusters based on their relevance to different control loops. To define such a relevance measure, we rely on a theoretical result from [KBT12]. It offers three stochastic correlation-based measures that indicate *usefulness* of incorporating a new sensor y_a or a new actuator u_a in a present system model, which is

considered as follows:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu_p(k) + \omega(k) \\ y_p(k) &= Cx(k) + Du_p(k) + v(k) \end{aligned} \tag{3.11}$$

In (3.11), x, u_p and y_p stand for states, present inputs and present outputs of the discrete time LTI system model with the input disturbance $\omega(t)$ and the measurement noise $v(t)$.

The three mentioned measures are employed as application-based routing metrics in order to extend the concept of *distance*. Consequently, membership in a specific cluster Λ_j is granted to a node if that node is closer to the cluster head than a specific threshold value d_j.

3.2.3.1 Addition of a new sensor

Addition of a new sensor is illustrated in Fig. 3.21.

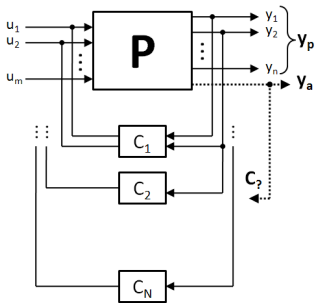

Figure 3.21: Block diagram of a PnP control system: The case of addition of a new sensor that should be connected to the relevant controllers

For a newly added sensor node, the following two complementary measures are proposed.

$$d^2_{U_p,y_a} = \frac{E[y_a(k) - \hat{y}_a(k|U_p^{k-1})]^2}{E[y_a(k) - E(y_a(k))]^2} \tag{3.12}$$

$$d^2_{U_pY_pY_a,y_a} = \frac{E[y_a(k) - \hat{y}_a(k|U_p^{k-1}, Y_p^{k-1}, Y_a^{k-1})]^2}{E[y_a(k) - E(y_a(k))]^2} \tag{3.13}$$

in which d^2 represents the correlation based distance and varies between 0 and 1. Subscripts $(.)_p$ and $(.)_a$ refer to present model and added device, respectively. $y(k)$ and $u(k)$ mean individual samples of a sensor's data and an actuator's command at time k, while Y^k and U^k indicate the set of all samples from the beginning up to and including time k. $E(.)$ stands for expected value operator over a finite number (N) of data samples which are treated as random variables [KBT12]. N is pre-defined in the sensor node. Superscript $\hat{(.)}$ stands for the least-squares estimation based on the available model.

With respect to the above mentioned definitions, interpretation of (3.12) and (3.13) is given in the following paragraph, assuming that: 1) the model of the present system is discrete-time linear time-invariant and 2) the new sensor provides sufficiently exciting data, and 3) a consistent un-biased least-squares estimation is given when $N \to \infty$.

The denominator in (3.12) and (3.13) is the variance of the data gathered by the new sensor, i.e. y_a. The numerator in (3.12) indicates how predictable the current $y_a(k)$ is if the commands of all actuators are known in the previous samples. If, according to the present model, none of the $k-1$ samples of all of the actuators have any tangible effect on the k^{th} sample of y_a, the following equation holds true.

$$E[y_a(k)|U_p^{k-1}] = E[y_a(k)] \tag{3.14}$$

Furthermore, if an unbiased estimation is assumed, we have $\hat{y}_a(k|U_p^{k-1}) = E[y_a(k)|U_p^{k-1}]$ which in combination with (3.14) results in the following expression:

$$\hat{y}_a(k|U_p^{k-1}) = E[y_a(k)] \tag{3.15}$$

Equation (3.15) means that the conditional least squares estimation of y_a is equal to its actual expected value. Therefore, the present model is good enough and the new measurement does not add any value to it. In this situation, $d^2_{U_p,y_a} = 1$, which should be read as: the new node is too far from the cluster head and cannot become a member, that is it is irrelevant to the control loop in question.

Equation (3.12) measures how much the additional sensor is affected by the present actuators in open loop. Nevertheless, this measure only reveals linear correlation. To look for nonlinear correlations, (3.12) should be modified according to the specific nonlinearity we are looking for. This is the easy step, but the difficult part is to perform nonlinear online incremental system identification to find $\hat{y}_a$. We do not consider this case in this paper.

The numerator in (3.13) measures how much the additional output could be controlled by the present actuators in closed loop. The interpretation is similar to (3.12), but this time the data from all of the sensors, including the new one, are also used in the least squares estimation, hence making it a more computationally intensive problem. Either (3.12) or (3.13) could be used in a given setting. Exploiting (3.12) is recommended in cases where y_a cannot be controlled independently of y_p [KBT12].

3.2.3.2 Addition of a new actuator

When a new actuator is added, the following measure is proposed.

$$d^2_{U_a y_p|U_p Y_p} = \frac{E[y_p(k) - \hat{y}_p(k|U_p^{k-1}, Y_p^{k-1}, U_a^{k-1})]^2}{E[y_p(k) - \hat{y}_p(k|U_p^{k-1}, Y_p^{k-1})]^2} \tag{3.16}$$

Equation (3.16) measure how much influence the additional actuator has on the present sensors in closed loop. If U_a^{k-1} does not have any effect on improving prediction of $\hat{y}_p$, then the prediction errors in the numerator and denominator will look alike and d will get its maximum value ≈ 1 On the other hand, if U_a^{k-1} is useful such that the prediction error in numerator is much less than that in denominator, then we have: $d \to 0$. Equation (3.16) should be interpreted similar to the previous measures with similar concerns. The same assumptions hold for a consistent estimation of $\hat{y}_p$. In practice, to provide a sufficiently exciting control signal, the actuator has to be driven by an external signal.

Remark 1: Latency in the communication network has a considerable impact on all of the introduced measures. But at the same time, it influences control performance too. Therefore, it is reasonable to consider this effect on evaluating usefulness of adding new sensors and actuators.

3.2.4 Routing Protocol Details

3.2.4.1 Network Layer Packet Format Details

To illustrate details of the protocol, the packet format shown in Fig. 3.22, is chosen for the Network layer.

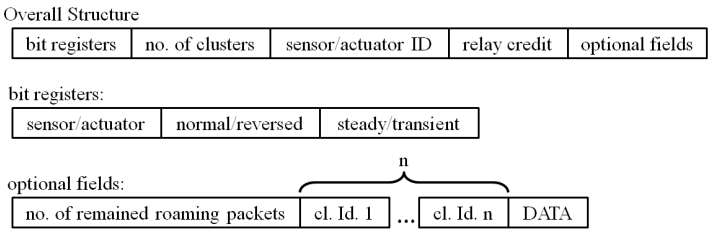

Figure 3.22: Structure of data packets at Network layer

The first field in the packet is a set of bit registers whose first bit determines whether the packet is sensor or actuator related. A sensor packet is generated in one of the sensors and should be routed to one or more controller nodes. An actuator packet is generated in a controller node and should be routed to an actuator node. The second bit indicates the direction of data. For sensor packets, "normal" means "from sensor to controller" and "reverse" means the other way around. The converse is true for actuator packets. The third bit register shows if the packet is sent in *steady* or *transient* operating mode. In transient mode, the packet contains an additional field, namely *number of remaining roaming packets*, which is listed among optional fields. The difference between these two modes and the function of roaming packets is explained later in **3.2.4.2**.

The second field stores the number of associated clusters. "Zero" in this field means that the packet is not associated with any cluster. If the packet is linked to $n > 0$ clusters, n additional fields are included in the packet, each of which containing the address of one of the associated clusters. For actuator packets, n could be either 0 or 1.

The next field is *Sensor/Actuator ID*. In sensor packets, it contains the address of the sensor node that has generated the packet. In actuator packets, it contains the address of the actuator node that the packet is destined to. Assignment of unique addresses to all of the nodes in the entire network is a prerequisite to our routing solution. The same demanding requirement exists in emerging standards, e.g. IETF RPL, by incorporating IPv6 as a worldwide addressing standard [NTWc].

The packet also contains a *relay credit* ≥ 0 which is explained in details, earlier in **3.2.2**. The last optional field is DATA which actually contains the application layer packet.

3.2.4.2 Cluster Formation: The role of Relay Credit

Here, we give a high level description of cluster formation. Assume that the controller nodes ($C_i, i = 1, ..., n$) are deployed as cluster heads. In a realistic scenario, each cluster head has a built-in model of the subsystem it is supposed to control. All of the initially deployed sensor and actuator nodes are already bound to their controllers. In other words, in the network setup phase, all of the nodes are aware of their cluster membership. As

a result, sensor nodes immediately start to function in their normal operating mode. See Fig. 3.23.

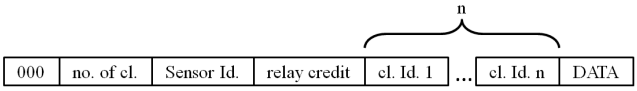

Figure 3.23: Structure of sensor packets at steady operation

Actuator nodes should start working in a safe mode and wait until they receive commands from the cluster head, i.e. the controller unit. The cluster head starts sending commands to the actuator as soon as it can devise the commands based on received sensor packets and the pre-programmed plant model. Command carrying packets are illustrated in Fig. 3.24.

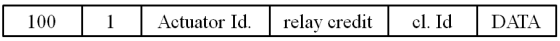

Figure 3.24: Structure of actuator packets at steady operation

In both above cases, the packets flood their pertinent clusters. Moreover, they propagate among the nodes of neighboring clusters into a certain depth defined by their relay credit.

Later on, when a new sensor pops up, it does not initially belong to any cluster and it is in transient operating mode. Thus, it publishes data in *roaming packets* as shown in Fig. 3.25.

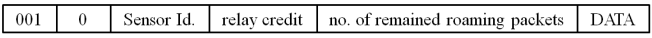

Figure 3.25: Structure of packets generated by a sensor when it is just turned on

In the beginning, the new sensor is not a member of any cluster. Hence, its data packet needs to spend relay credit to get around. When a roaming packet arrives at a neighboring node that is operating in *Steady* mode, it inherits all of the cluster tags of that node – meaning that the $cl.Id.$ fields of the packet are refreshed. This action is performed only if *relay credit* > 0. If either the neighboring node is in *Transient* mode or *relay credit*$= 0$, cluster tags of the packet remain unchanged.

In the steady mode, embedding a cluster tag into a packet gives it the right to freely flood in that cluster without spending relay credit, but it is not the case in transient mode in which relay credit is constantly spent for every transmission. Therefore, embedding cluster tags into roaming packets does not give them a free pass. On the other hand, running out of relay credit is not the stop criterion when retransmitting a packet in transient mode. It just kills its ability to inherit new cluster tags. At the end, a roaming packet floods into the clusters that it managed to enter before running out of relay credit.

Example 1: Fig. 3.26 depicts an example when a new sensor is placed amongst nodes of Λ_1. However, some of its roaming packets could also reach the borders of Λ_2 before consuming all of their relay credit. Thus, presence of the new sensor is advertised through the union of nodes of Λ_1, Λ_2, and in the *relay credit* > 0 zone. Based on the above setting,

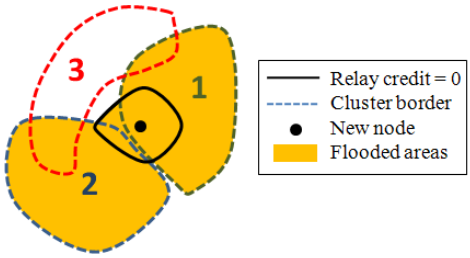

Figure 3.26: Effect of *relay credit* when a new node is joined

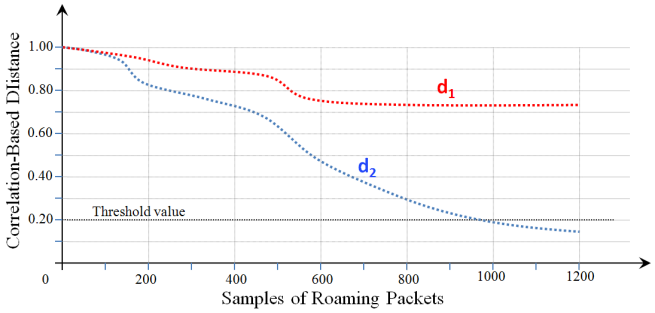

Figure 3.27: Development of awareness measure (3.12) in C_1 and C_2 during identification process

C_1 and C_2 start calculating (3.12), or (3.13), or both. Note that, C_3 might also receive the roaming packets of the new sensor if it is placed in $\Lambda_3 \bigcap \Lambda_1$ or $\Lambda_3 \bigcap \Lambda_2$, but it will not calculate (3.12) or (3.13) because Λ_3 is not listed among clusters in the packets.

Fig. 3.27 shows how the correlation-based distance measures (3.12) or (3.13) might develop over time in C_1 and C_2. It is assumed that the new sensor generates 1200 roaming packets, i.e. one per sample time.

After collecting sufficient samples at the cluster heads, each C_i decides whether the new sensor should become a member of their cluster or not. In Fig. 3.27, C_2 concludes that the new sensor is relevant well before the roaming packets are discontinued, but C_1 finds the new sensor irrelevant to the control performance of its internal model. C_2 continues the joining process by sending a *join* request to the new sensor. The data packet which contains the join request is in the following form, shown in Fig. 3.28.

010	1	Sensor Id.	relay credit	cl. Id. 1

Figure 3.28: Join request from cluster head to a sensor

Note that this packet is issued in ”steady operating mode”, which means no relay credit is deducted unless it leaves the source cluster, Λ_2 in our example. This mechanism is useful when the new node is only reachable via nodes of other clusters.

After sending all roaming packets, 1200 in our example, the new sensor applies re-

ceived join requests. Join requests arrive at the sensor node asynchronously. Therefore, the node should continually accept join requests, at least until a pre-defined time. If no join request is received, the sensor starts another round of generating roaming packets.

If a group of sensors are deployed simultaneously, the ones which have other cluster members in their vicinity will find their clusters earlier. The others that do not have any neighbor operating in steady mode, may not inherit cluster tags, hence their flooding domain is limited to the *relay credit* > 0 zone.

The above procedure is slightly different for a new actuator. Each newly turned on actuator applies a pre-specified control sequence for the purpose of sufficiently exciting the plant and creating measureable outcomes. However, it should be sufficiently cautious not to disturb the plant too much. Simultaneously, it publishes roaming packets as shown in Fig. 3.29, which contain current value of the actuator output.

111	0	Actuator Id.	relay credit	no. of remained roaming packets	DATA

Figure 3.29: Packets generated by an actuator when it is just turned on

The cluster heads which receive these packets, use the DATA field in evaluating (3.16) similar to what was shown in Fig. 3.27. When the roaming packets are discontinued – meaning that the actuator is waiting for the decision – each cluster head returns the calculated *usefulness measure* to the actuator by packets shown in Fig. 3.30.

100	1	Actuator Id.	relay credit	cl. Id. 1	DATA

Figure 3.30: Packets that return usefulness of utilizing an actuator in a cluster

The actuator waits for a certain time to receive evaluation results from all involved cluster heads. Then it compares the received usefulness values, which are embedded in DATA fields, and selects the cluster that has returned the highest value. If at least one evaluation result is received, the actuator chooses its own cluster and sends a join request to that cluster as illustrated in Fig. 3.31. If no evaluated usefulness measure is received, the above procedure starts from the beginning.

110	1	Actuator Id.	relay credit	cl. Id. 1

Figure 3.31: Join request from an actuator to a cluster head

Note that, when a new sensor is added, it receives individual join requests from controllers. But when a new actuator is added, it sends the join request to a single controller.

3.2.4.3 Performance Related Remarks

Remark 1: Unlike other clustered flooding-based routing protocols whose acceptable performance depend heavily on the optimal choice of the number of clusters and the thoughtful selection of cluster heads [HCB02, YF04], these parameters are pre-defined in our protocol because all of the controller units ($C_j, j = 1...n$) and only the controller units are cluster heads. Moreover, the controller nodes are fixed in the whole lifetime of the network.

Remark 2: Another issue is the influence of lower layer protocols on performance of the routing protocol. Unlike [HCB02], that has utilized a TDMA-based MAC for the sake of energy-efficient collision-free transmissions, the MAC layer in our system cannot accommodate a deterministic reservation-based protocol. It is mainly due to the constrained coverage range of nodes which does not guarantee existence of direct links between cluster heads and every member of the cluster to schedule a frame-based MAC. After all, it is a prerequisite for framed MACs that all of nodes can be accessed from a single base station for scheduling purposes. Otherwise, many time frames must be kept unused and reserved for future extension, as in Time Synchronized Mesh Protocol (TSMP) [PD08].

Our routing solution may be built either on a contention-based or a preamble-sampling MAC which are inferior to deterministic MACs in terms of energy efficiency and end-to-end latency if data transmission among nodes is frequent. Therefore, the main benefit of our routing solution is to pick the members of each cluster prudently such that it results in the minimum number of nodes in a cluster, and more efficient flooding in clusters.

3.2.5 Simulation results

In this section, we evaluate effects of the e2e latency and jitter on our data-centric clustering method via a simulation example. The simple version of this example without considering communication constraints is already presented in [KBT12]. Our purpose is to find out how the system model and the network structure will develop if a new sensor joins the network in presence of latency and jitter. The original system model is:

$$\begin{bmatrix} x_1(t+1) \\ x_2(t+1) \end{bmatrix} = \begin{bmatrix} 0.9048 & -0.0090 \\ 0.0090 & 0.9048 \end{bmatrix} . \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0.0952 & 0.0947 \\ 0.00047 & 0.0956 \end{bmatrix} . u_p(t) + \begin{bmatrix} w_1(t) \\ w_2(t) \end{bmatrix} \tag{3.17}$$

$$y_p(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} . \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 & 0.1 \end{bmatrix} . u_p(t) + v(t) \tag{3.18}$$

with noise correlations:

$$R_w = \begin{bmatrix} 0.0100 & 0.0050 \\ 0.0050 & 0.0100 \end{bmatrix}, R_v = 0.01, R_{wv} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{3.19}$$

The original system is intentionally chosen such that $x_2(t)$ has poor observability. This is implied by the large condition number (≈ 201) of the observability matrix. In the simulation example in [KBT12], an aggresive LQ state feedback controller (3.20) is used in combination with a kalman filter whose gain and innovation covariance are given in (3.21).

$$u(t) = - \begin{bmatrix} 2.3953 & -1.2122 \\ 1.2138 & 2.4299 \end{bmatrix} . \hat{x}(t) \tag{3.20}$$

$$K_p = \begin{bmatrix} 0.5381 \\ 0.2597 \end{bmatrix}, \quad R_{ep} = 0.0248 \tag{3.21}$$

It is also illustrated in [KBT12], how a new output measurement signal $y_2(t)$, as described in (3.22), will contribute to improvement of: 1) observability of the second state, and 2) the total cost of the unaltered LQ controller by just updating the kalman filter.

$$y_2(t) = \begin{bmatrix} 1 & 1 \end{bmatrix} . \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0.1 & 0.1 \end{bmatrix} . \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix} + v_a(t) \tag{3.22}$$

The block diagram of the system is shown in Fig. 3.32. It is a special case of the general problem setup in Fig. 3.21, in which there is only one controller and we are interested to see if the added remote sensor will improve control performance. If there were more than one controller, the effect on control performance of each controller would imply whether the new sensor should stay connected to that controller, and hence become a member of its cluster.

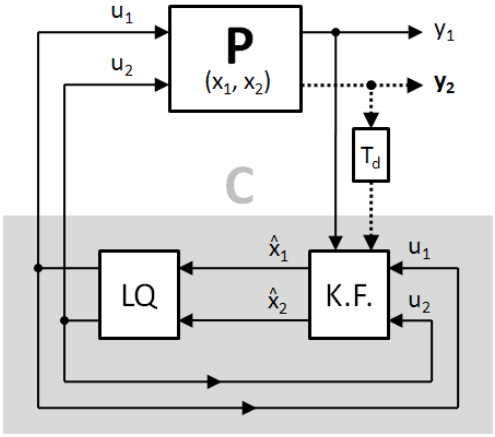

Figure 3.32: A simple PnP control system which benefits from addition of a new sensor whose data is delayed T_d due to network latency

Assuming the state and control input cost functions are defined as in (3.23) and (3.24), and taking no communication constraints into account, i.e. $T_d = 0$, improvements of the incremental model compared to the original model are shown in Tab. 3.3 for the sake of completeness.

$$\begin{aligned} J_x &= \frac{1}{N} \sum_{t=1}^{N} x(t)^T Q_x x(t) \\ &= \frac{1}{N} \sum_{t=1}^{N} (x_1(t)^2 + x_2(t)^2) \end{aligned} \tag{3.23}$$

$$J_u = \frac{1}{N}\sum_{t=1}^{N} u(t)^T Q_u u(t)$$
$$= 0.05\frac{1}{N}\sum_{t=1}^{N}(u_1(t)^2 + u_2(t)^2) \tag{3.24}$$

Table 3.3: Modification of the state cost, the control cost, and the total cost of the updated system compared to the original system [KBT12]

	J_x	J_u	$J = J_x + J_u$
Updated/Original	0.5851	1.6265	0.6340

We have extended the above mentioned example by considering the original model as a lumped model without any communication constraint, but the added sensor as a remote sensor, whose data has to be transferred via a network. The network will induce a latency (T_d) on the newly added measured data as expressed in (3.25).

$$y_a(t) = y_2(t - T_d) \tag{3.25}$$

Assuming stability of the routes through the network, which is mainly provided by the channel hopping and channel agility mechanisms recommended in Section II, the induced latency dynamics is affected only by the MAC protocol. Since either the IEEE802.15.4e or the IEEE802.11e were already chosen as the MAC layer solution, two scenarios were possible. In the first one, a relatively large latency without jitter was imposed to the data from the added sensor. It is meant to be understood as the abstracted effect of choosing TSCH in the PA application domain of the IEEE802.15.4e. In the second scenario, associated with the EDCA mode of the IEEE802.11e, relatively small latencies were applied which were subject to non-zero jitter with uniformly distributed values.

Fig. 3.33 shows simulation results for the first scenario. Among the cost functions in Tab. 3.3, only the state cost is of interest because the control cost is not improved even in the original example without any latency. Actually, any improvement in the state cost is achieved by employing larger control signals. The total cost is also a summation which can easily be modified by changing Q_x and Q_u in (3.24) and (3.23). That is why only the state cost is investigated and illustrated in Fig. 3.33.

As Fig. 3.33 shows, control performance degrades as latency increases. This continues until a certain amount of latency after which, the identification process fails and the use of an incorrect model results in very large oscillations, i.e. stochastic behavior, in the state cost. This means that the parameters of the updated model are not converged anymore.

In this example, the largest admissible latency is 9 times larger than one sample time. It is achieved when $jitter = 0$ and the parameters in the system identification problem can still converge in presence of the maximum tolerable latency. It can be seen that for a jitter-free network, improvement on the state cost varies from 41.5% for a latency-free network to 21.2% when latency equals to 9 sample times. Moreover, it is assumed that

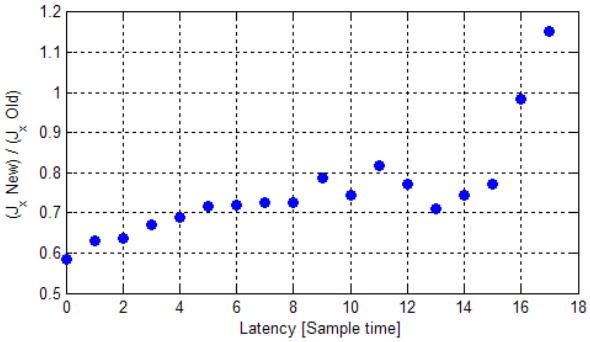

Figure 3.33: Influence of jitter-free latency on relative state cost

the network has 100% reliable links and a completely deterministic MAC protocol such as the TSCH mode of the IEEE802.15.4e.

Fig. 3.34 adds the effect of latency variation on the relative state cost. Each curve in this figure represents a stochastic specification for jitter. The simulation is done on the assumption that each roaming packet from the newly joined sensor is subject to a random latency which is a multiple of the sampling rate and is uniformly distributed in the interval $[Latency_{min}, Latency_{min} + 2 \times Jitter]$. Each curve is produced by increasing $Latency_{min}$ until numerical simulations become invalid, as explained above.

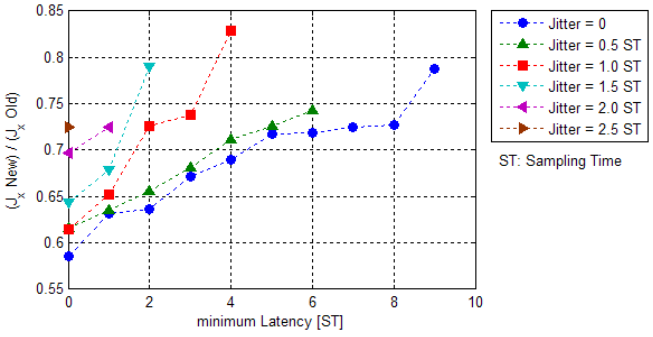

Figure 3.34: Influence of latency and jitter on relative state cost

Assuming that the average latency in all of the above cases equals to $Latency_{min} + Jitter$, the bar plot in Fig. 3.35 shows how jitter reduces the maximum tolerable average latency. Each bar in this figure is plotted based on the end point of individual curves in Fig. 3.34. It can be seen how deterministic behavior contributes to pushing the limits of the tolerable latency forward.

The simulation results confirm that wireless PnP control systems benefit most from completely deterministic solutions at the MAC layer. They also show how a control oriented scalar measure can act as a routing metric and contribute to clustering of the nodes of a network in a flooding-based routing algorithm. The proposed routing metric makes use of model-based correlation estimation between the new nodes and the existing model of the system.

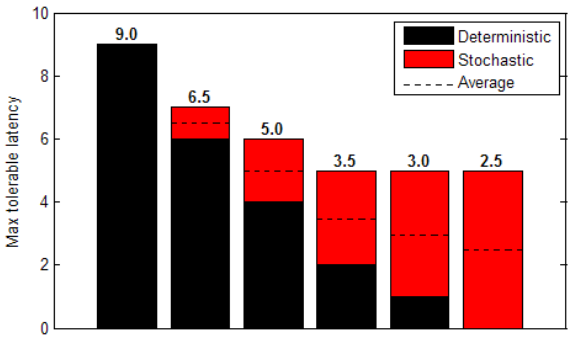

Figure 3.35: Maximum tolerable average latency

4 | Summary of Contributions

The contributions of this thesis are in two distinct areas which are: 1) signal propagation in underground and confined areas, and 2) access and networking protocols in WPnP control systems. Therefore, they are submitted and published in conferences and journals of different communities. The rest of this chapter enumerates the contributions in both areas on a chronological order.

4.1 A method for signal propagation in underground wireless networks

Having focused on the district heating system case study, all of the known signal propagation methods that could be used to create communication links between two underground nodes with a known physical distance range were investigated. Eventually, Magnetic Induction was chosen as the most suitable option. The results are documented in Paper A, Paper B, and Paper C. These three papers contain the following contributions:

1. As the starting point in Paper A, physical distance scattering among nodes of a district heating system was found based on a real scenario. Then, the idea of using district heating pipes as the communication medium was pursued. The idea was motivated by the fact that all of the nodes are connected via steel pipes with carefully welded joints. Unlike potable water distribution network, there are a number of special requirements for district heating pipes that should tolerate very high pressures and temperatures. Such requirements including: 1) material specification, 2) geometric constraints on joints, and 3) the acceptable welding procedure, could facilitate building a communication framework which relies on pipes. The main contribution of Paper A was to exploit magnetic properties of steel pipes to propagate signals among the nodes. The results were achieved by finite element simulations.

2. Finite element simulations were done on steel pipes to calculate mutual inductance between transceiver coils. Simulation results foresaw applicability of the method in Paper A. However, those promising results could not be verified by experiments that were done on district heating pipes in two adjacent university buildings. Although the experiments, which were a failure, are reported neither in any paper nor earlier in the thesis, they contribute to not pursuing the idea of using steel pipes as magnetic signal conductors. The reason for failure of experiments was to rely on too simplified assumptions in the simulations of Paper A.

3. Thereafter, the work continued by doing a thorough survey of other signal propagation methods. Although the most complete survey appeared later in Paper C, the favored method of Magnetic Induction is chosen and introduced first in Paper B. Paper B, mainly deals with optimum coil antenna design for magneto-inductive communication systems via simulations. It verifies feasibility of signal propagation through non-ferromagnetic materials like air and soil which are of interest in the district heating system case study. The results in Paper B are based on the exact analytical formula of the magnetic field that is created by a multi-turn coil of wire in the space. The optimum number of turns and dimensions of transmitter and receiver coils were found assuming pre-defined wire length and gauge. The assumption on wire gauge and length originates from major considerations of price and weight of the antennas. It is shown that for the transmitter antenna, a multi-layered coil with the maximum possible length and the minimum mean radius is preferred. For the receiver antenna, on the other hand, the largest possible mean radius with a circular cross section is optimum.

4. Finally, the simulation based results of Paper B, are accompanied by experimental verification in Paper C.

   The most viable results in the first part of the thesis are reported in Paper C. It starts by highlighting differences between the current and the new piping layout of district heating systems, advantages of the new layout, and complexities of the new control system. It reviews all of the signal propagation methods that were investigated in this PhD project. They include: Through the soil radio frequency (RF) electromagnetic (EM) waves, Combined underground above-ground exploitation of RF EM waves, Pipelines as EM waveguides, Acoustic waves through the pipes, Acoustic waves through the in-pipe water, Electrical conductivity of pipes, Power Line Communications, Cell phone infrastructure, Magnetic induction through steel pipes, and Air-core magnetic induction.

   Paper C continues by reproducing the main result of Paper B on coil antenna design. Furthermore, it completes the simulation based results of Paper B by presenting experimental verification. A pair of coil antennas was used in an office environment to conduct the tests. Coils were wound based on optimum coil antenna recommendations. Paper C is concluded by an affirmative answer to the question of feasibility of underground communication by using magnetic induction, in a problem setting similar to the case study of district heating system.

Last but not least, though at the time of reviewing the aforementioned methods for signal propagation in Paper C, air-core magnetic induction was chosen as the best solution, the best candidate may change due to technology improvements in other areas. Especially cell phone infrastructure and power line communications (PLC) are rapidly evolving that might take over MI sooner or later. New research areas like Smart Grid bring more stamina to specific areas like PLC. Therefore, when it comes to practice and implementation, one should be aware of the latest changes in these two specific areas. Maybe the race between competing standards for taking control over PLC resources will be finished under the new ruling of Smart Grid in near future.

4.2 A routing solution for wireless plug and play control systems

A large number of access and networking protocols have been proposed for wireless sensor networks based on a variety of topologies. The inherent properties of the case studies of P^3C, however, have diminished usefulness of many of them by imposing a multi-hop structure and P2P traffic pattern. In the beginning of this research, it was intended to only focus on routing protocols in WSNs assuming that the main source of latency in wireless industrial networks is the multi-hop structure of the network. Later, it turned out that besides the latency due to multi-hop structure, there is another underlying difference between wireline and wireless industrial networks and that is the additional considerations in designing the MAC protocol. Therefore, MAC protocols were also studied. The main results on MAC and routing protocols in WPnP control systems are presented in Paper D and Paper E and highlighted as follows.

1. Paper E presents a survey on potentially viable solutions in the MAC layer. The key point in selecting the relevant MAC protocols was their *deterministic* performance. The contribution of the thesis in this area is limited to aggregation and comparison of several protocols based on findings of other researchers. Therefore, it is only presented in the state of the art section of the thesis and Paper E. This part concludes by recommendations in favor of the IEEE802.15.4e which is a developing standard.

2. The same approach of surveying relevant protocols was considered in dealing with the routing problem. Available protocols are categorized in a new way based on dividing the problem of routing into two tasks, namely: structuring the network and proposing a routing algorithm. Then, the assumptions on predominant data traffic pattern and the networking topology, which were suggested earlier, were utilized to undermine all *gradient-based* routing algorithms, and hence promote the use of *flooding-based* algorithms instead. This result is reported in Paper D in details.

3. Every practical flooding-based routing algorithm should be accompanied by a an efficient network structuring method to constrain the flooding domain of each packet and alleviate routing overhead. Without a proper structuring method, pure flooding imposes too many unnecessary retransmissions.

   Although many network structuring methods are proposed in the literature, none of them were directly influenced by the top layer application. Even those that claimed to make such links were only designed to satisfy general requirements like providing balance in residual energy of all nodes. Nevertheless, this thesis has offered a real application-based routing metric. It has gone one step further from traditional designs in which only effects of communication constraints on the control application were taken into account.

   The proposed routing metric is introduced first in Paper D as a means of clustering nodes based on their relevance to different control loops. The basic idea is to form clusters while the procedure of incremental system identification is in progress. As soon as the control application in a controller node, which is also a cluster head, finds out how much a new sensor / actuator contributes to the control system performance, it decides whether to include that sensor/actuator node into its own cluster. Paper D also presents a conceptual example, but the functionality of that metric was confirmed later in the simulations in Paper E.

4. As the first step towards implementation of the proposed routing solution, a specific packet format is presented for the network layer in details in Paper D. Step by step procedures are scrutinized from the moment when new sensors and actuators pop up until they join certain clusters. Moreover, a new concept called *relay credit* is introduced in order to preserve connectivity of the nodes that are relevant in view point of the control application, but do not consist a connected segment.

5 | Conclusions and Future Works

This chapter, similar to the previous ones, is divided into two independent sections.

5.1 Conclusions and future works on physical communications

Due to the necessity of covering all of the case studies of P^3C and because of the outcome of the case study survey, it was required to specifically look into the District Heating Case study. The totally different signal propagation environment made further investigations essential to assess feasibility of making communication links among nodes of the network.

Among the examined technologies, power line communications and magnetic induction turned out to be the most viable options. Power line communications has already found many applications that are competing for the limited resources on power lines. Although the large variety of applications has put off standardization activities, the situation may change in near future if more urban utilities are integrated into emerging Smart Grids with quickly converging standards that are under progress yet.

Magnetic induction, on the other hand, has only been used in very specific domains so far, with high power consumption and extended ranges. It is anticipated in this thesis that this technology can be used with mild power requirements in underground urban utilities in relatively short range multi-hop networks. The bottom line is to systematically quantify effects of ferromagnetic materials in the vicinity of transceivers. Furthermore, since the operating frequency is very low compared to other RF solutions, the available bandwidth is limited. This makes the solution not interesting for broadband applications. Therefore, magnetic induction is exclusively useful for sensor networks which require a small amount of bandwidth with guarantees on availability of the channel as soon as it is needed.

Supported by the promising simulation and small scale experimental results, an immediate future work in this regard is to design and run experiments on full scale district heating pipes. Such tests could be started by only one pair of transceivers located close together and eventually extend the range in order to verify the results on the real framework.

Another extension of the work is to inspect whether only one coil antenna at each node can be used both as the transmitter and the receiver antennas. Two way links are needed if the MAC protocol is based on information exchange, e.g. in token ring or in contention based MAC protocols. Therefore, the problem is to see if a compromise between the ideal Tx and Rx specifications can be made to make a coil operating good enough both as a Tx

and a Rx antenna in a half-duplex scheme.

5.2 Conclusions and future works on access and routing protocols

Assuming low bandwidth requirements of WPnP control systems and with emphasis on latency and jitter, utilizing a clustered flooding-based routing algorithm on top of a deterministic MAC protocol are found to be the viable solution in this thesis. Unlike the general trend of ever-increasing interests in the convergecast traffic pattern and gradient-based routing solutions, it was stated that a true P2P architecture is preferred if closed loop reconfigurable control loops are to be realized. As a consequence, flooding-based routing algorithms were found to be superior compared to their gradient-based counterparts.

To mitigate the classical shortcomings of flooding-based routing, a novel data-centric clustering mechanism is proposed to constrain the domain of flooding. The interesting issue in this part is the method to form and maintain these clusters which directly employs data from the top layer control application. This also provides a clear link between WP1 and WP2 of the P^3C project and involves control system data in the network protocol design.

Influence of the proposed routing protocol on the networked control system is evaluated via simulations. Although the choice of the MAC protocol is not directly seen in the simulations, the possible effects of the IEEE802.15.4e and the IEEE802.11e MAC protocols are estimated by applying a range of latencies and spanning parameters of distribution of jitter.

Since functionality of the routing mechanism is only assessed in simulations with abstracted MAC protocol effects, and by applying the proposed packet formats, an immediate future work could be to implement the routing algorithm on top of a network simulator either with the IEEE802.15.4e or the IEEE802.11e as chosen MAC protocols. The simulation framework should accommodate top level application since the routing algorithm requires interacting with the application in order to function.

References

[APM05] Ian F. Akyildiz, Dario Pompili, and Tommaso Melodia. Underwater acoustic sensor networks: research challenges. *Elsevier Ad Hoc Networks*, 3(3):257–279, May 2005.

[Arr81] R.A. Arrington. Load management and automatic meter reading through the use of power line carrier. *Electric Power Systems Research*, 4(2):85–104, April 1981.

[AS06] I.F. Akyildiz and E.P. Stuntebeck. Wireless underground sensor networks: Research challenges. *AdHoc Networks Journal*, 4(6):669–686, November 2006.

[ASV09] I.F. Akyildiz, Z. Sun, and M.C. Vuran. Signal propagation techniques for wireless underground communication networks. *Physical Communication*, 2(3):167–183, September 2009.

[Bal89] Constantine A. Balanis. *Advanced engineering electromagnetics*. John Wiley & Sons Inc., New York, 1989.

[Ban04] R. Bansal. Near-field magnetic communication. *IEEE Antennas and Propagation Magazine*, 46(2):114–115, April 2004.

[Bat20] R.R. Batcher. Loop antenna for submarines. *Wireless age*, 7:28, 1920.

[BBG05] L. Blazevic, J.Y. Le Boudec, and S. Giordano. A location-based routing method for mobile ad hoc networks. *IEEE Transactions on Mobile Computing*, 4(2):97–110, March-April 2005.

[BCCF02] Flaminio Borgonovo, Antonio Capone, Matteo Cesana, and Luigi Fratta. RR-ALOHA, a Reliable R-ALOHA broadcast channel for ad-hoc inter-vehicle communication networks. In *Med-Hoc-Net*, 2002.

[BCCF04] Flaminio Borgonovo, Antonio Capone, Matteo Cesana, and Luigi Fratta. ADHOC MAC: a new MAC architecture for ad hoc networks providing efficient and reliable point-to-point and broadcast services. *ACM Wireless Networks*, 10(4):359–366, July 2004.

[BDWL10] Abdelmalik Bachir, Mischa Dohler, Thomas Watteyne, and Kin K. Leung. MAC essentials for wireless sensor networks. *IEEE Communications Surveys & Tutorials*, 12(2):222–248, November 2010.

[BT80] J. Bhagwan and F. Trofimenkoff. Mutual inductance of coils on a cylinder of infinite length. *IEEE Transactions on Magnetics*, 16(2):477–479, March 1980.

[Bun01] C. Bunszel. Magnetic induction: a low-power wireless alternative. *Defense Electronics, Formerly: RFDesign*, 24(11):78–80, November 2001.

[BVR+04] Benny Bohm, Niels Kristian Vejen, Jan Rasmussen, Niels Bidstrup, Kurt Paaske Christensen, Finn Bruus, and Halldor Kristjansson. EFP-2001 Fjernvarmeforsyning af lavenergiomraader. Technical report, Energistyrelsen, March 2004. In Danish.

[CJ03] T. Clausen and P. Jacquet. Optimized link state routing protocol (olsr), October 2003.

[Cla98] David Clark. Powerline communications: finally ready for prime time? *IEEE Internet Computing*, 2(1):10–11, January 1998.

[CP10] I. Chakeres and C. Perkins. Dynamic manet on-demand (dymo) routing, July 2010.

[CWLY06] Min Chen, Xiaodan Wang, Victor C. M. Leung, and Yong Yuan. Virtual coordinates based routing in wireless sensor networks. *Sensor Letters*, 4(3):325–330, September 2006.

[DHS] Danfoss heating book - 8 steps to control of heating systems. Available from: `http://heating.danfoss.com/Content/61078BB5-1E17-4CA2-AC49-8A7CAC2BA687\_MNU17424997\_SIT54.html`.

[DUHER85] Myron C. Dobson, Fawwaz T. Ulaby, Martti T. Hallikainen, and Mohamed A. El-Rayes. Microwave dielectric behavior of wet soil – part II: Dielectric mixing models. *IEEE Transactions on Geoscience and remote sensing*, GE-23(1):35–46, January 1985.

[ELSV04] Mustafa Ergen, Duke Lee, Raja Sengupta, and Pravin Varaiya. WTRP – Wireless Token Ring Protocol. *IEEE Transactions on Vehicular Technology*, 53(6):1863–1881, November 2004.

[EMS+05] K.T. Erickson, A. Miller, E.K. Stanek, C.H. Wu, and Sh. Dunn-Norman. Pipelines as communication network links. Technical report, University of Missouri-Rolla, March 2005.

[Erg02] Mustafa Ergen. WTRP-Wireless Token Ring Protocol. Master's thesis, University of California, Berkeley, 2002.

[FS06] Hannes Frey and Ivan Stojmenovic. On delivery guarantees of face and combined greedy-face routing in ad hoc and sensor networks. In *Proceedings of the 12th annual international conference on Mobile computing and networking*. ACM, 2006.

[GSW11] Stefano Galli, Anna Scaglione, and Zhifang Wang. For the grid and through the grid: The role of power line communications in the smart grid. *Proceedings of the IEEE*, 99(6):998–1027, June 2011.

[HCB02] W.B. Heinzelman, A.P. Chandrakasan, and H. Balakrishnan. An application-specific protocol architecture for wireless microsensor networks. *IEEE Transactions on Wireless Communications*, 1(4):660–670, October 2002.

[HEB08] M. Hamdi, N. Essaddi, and N. Boudriga. Energy-efficient routing in wireless sensor networks using probabilistic strategies. In *IEEE Wireless Communications and Networking Conference (WCNC)*, pages 2567–2572, Las Vegas, NV, USA, April 2008.

[HUD+85] M.T. Hallikanien, F.T. Ulaby, M.C. Dobson, M.A. El-Rayes, and L.K. Wu. Microwave dielectric behavior of wet soil – part I: Empirical models and experimental observations. *IEEE Transactions on Geoscience and Remote Sensing*, GE-23(1):25–34, January 1985.

[IN99] Tomasz Imielinski and Julio C. Navas. Gps-based geographic addressing, routing, and resource discovery. *Communications of the ACM*, 42(4):86–92, 1999.

[JMB01] David B. Johnson, David A. Maltz, and Josh Broch. *DSR: The Dynamic Source Routing Protocol for Multi-Hop Wireless Ad Hoc Networks*, chapter 5, pages 139–172. Addison-Wesley, 2001.

[Kal07] C. Skovmose Kallese. Simulation of a district heating system with a new network structure. Technical report, Grundfos A/S, Bjerringbro, Denmark, February 2007.

[KBT12] Torben Knudsen, Jan Bendtsen, and Klaus Trangbaek. Awareness and its use in incremental data driven modelling for Plug and Play Process Control. *European Journal of Control*, 18(1):xx–yy, February 2012.

[KHB02] Joanna Kulik, Wendi Heinzelman, and Hari Balakrishnan. Negotiation-based protocols for disseminating information in wireless sensor networks. *Wireless Networks*, 8(2):169–185, March-May 2002.

[KMH+08] P. Kyosti, J. Meinila, L. Hentila, X. Zhao, T. Jamsa, Ch. Schneider, M. Narandzic, M. Milojevic, A. Hong, J. Ylitalo, V.M. Holappa, M. Alatossava, R. Bultitude, Y. de Jong, and T. Rautiainen. WINNER II channel models, part I, D1.1.2 V1.2, 2 2008. Available from: `http://www.ist-winner.org/WINNER2-Deliverables/D1.1.2.zip`.

[Kok06] George Kokossalakis. *Acoustic data communication system for in-pipe wireless sensor networks*. PhD thesis, Massachusetts Institute of Technology, February 2006.

[Kon05] Antonios Kondis. Acoustical wave propagation in buried water filled pipes. Master's thesis, Massachusetts Institute of Technology, February 2005.

[Kot88] Oleg Kotlyar. Universal mud pulse telemetry, September 1988.

[Kun07] Kevin Kuns. Calculation of magnetic field inside plasma chamber. Technical report, UCLA, August 2007.

[Lan96] Hkan Lans, 1996.

[LCL03] R. Long, P. Cawley, and M. Lowe. Acoustic wave propagation in buried iron water pipes. In *Proceedings A: Mathematical, Physical & Engineering Sciences*, volume 459, pages 2749–2770. The Royal Society, November 2003.

[Lee01] Duke Lee. Wireless token ring protocol. Master's thesis, University of California, Berkeley, 2001.

[Loc01] Evan Locke. Switchable transceiver antenna, December 2001.

[Loc04] Gordon Evan Locke. Apparatus and method for continuous variable reactive impedance control, April 2004.

[Max73] James Clerk Maxwell. *A treatise on electricity and magnetism*, volume 2 of *Clarendon press*. Oxford, 1873.

[Moo51] R.K. Moore. *Theory of radio communication between submerged submarines*. PhD thesis, Cornell University, 1951.

[MPD10] Soroush Afkhami Meybodi, Pablo Pardo, and Mischa Dohler. Magneto-inductive communication among pumps in a district heating system. In *the* 9^{th} *International Symposium on Antennas, Propagation and EM Theory*, pages 375–378, Guangzhou, China, December 2010.

[NB85] Munir Hassan Nayfeh and Morton K. Brussel. *Electricity and magnetism*. John Wiley & Sons Inc., New York, 1985.

[NTWa] Available from: `http://tools.ietf.org/wg/roll/`.

[NTWb] IETF ROLL: Industrial routing requirements. Available from: `http://tools.ietf.org/html/rfc5673`.

[NTWc] IETF ROLL: IPv6 Routing Protocol, draft standard. Available from: `http://tools.ietf.org/wg/roll/draft-ietf-roll-rpl/`.

[NTWd] IETF ROLL: P2P traffic support. Available from: `http://tools.ietf.org/wg/roll/draft-ietf-roll-p2p-rpl/`.

[NTWe] IETF ROLL: routing metrics used for path calculation in low power and lossy networks. Available from: `http://tools.ietf.org/wg/roll/draft-ietf-roll-routing-metrics/`.

[Ols02] Robert G. Olsen. Technical considerations for wideband powerline communication – a summary. In *IEEE Power Engineering Society Summer Meeting*, volume 3, pages 1186–1191, Chicago, IL, USA, July 2002.

[P3C06] Plug and play process control, research program of the danish research council for technology and production sciences, 2006. Available from: `http://www.control.aau.dk/plugandplay`.

[PBRD03] C. Perkins, E. Belding-Royer, and S. Das. Ad hoc on-demand distance vector (aodv) routing, July 2003.

[PCB07] Vincent Palermo, Patrick J. Cobler, and Neal R. Butler. Inductive communication system and method, August 2007.

[PD08] Kristofer S. J. Pister and Lance Doherty. TSMP: time synchronized mesh protocol. In *Proceedings of the IASTED International Symposium on Distributed Sensor Networks*, November 2008.

[PK09] Claudio De Persis and Carsten S. Kallesoe. Pressure regulation in nonlinear hydraulic networks by positive controls. In *European Control Conference*, pages 4102–4107, Budapest, Hungary, August 2009.

[Pop01] G.M. Popov. A method of calculating the complex magnetic permeability of a ferromagnetic rod. *Springer Measurement Techniques*, 44(5):522–528, May 2001.

[Pop03] G.M. Popov. A magnetic circuit in the form of an extended ferromagnetic rod magnetized by a coil of arbitrary length. *Springer Measurement Techniques*, 46(4):377–382, April 2003.

[RVD06] David Reagor and Jose Vasquez-Dominguez. Through-the-earth radio, December 2006.

[SA09] Z. Sun and I.F. Akyildiz. Underground wireless communication using magnetic induction. In *In Proc. IEEE ICC*, Dresden, Germany, June 2009.

[SF06] Hirohisa Sakuma and Jun Fujiwara. Acoustic communication device and acoustic signal communication method, April 2006.

[SGW99] John Sojdehei, Felipe Garcia, and Robert Woodall. Magneto-inductive seismic fence, October 1999.

[SNU06] H. Sakuma, K. Nakamura, and S. Ueha. Two-way communication over gas pipe-line using multicarrier modulated sound waves with cyclic frequency shifting. *Acoustical Science and Technology, The Acoustical Society of Japan*, 27(4):225–232, July 2006.

[SR02] R.C. Shah and J.M. Rabaey. Energy aware routing for low energy ad hoc sensor networks. In *IEEE Wireless Communications and Networking Conference (WCNC)*, pages 17–21, Orlando, FL, USA, March 2002.

[Sto08] Milica Stojanovic. Underwater acoustic communications: Design considerations on the physical layer. In *Wireless on Demand Network Systems and Services (WONS)*, pages 1–10, Garmisch-Partenkirchen, January 2008. IEEE.

[SWD01] J.J. Sojdehei, P.N. Wrathall, and D.F. Dinn. Magneto-inductive (MI) communications. In *Oceans, MTS/IEEE Conference and Exhibition*, pages 513–519, Honolulu, HI, USA, November 2001.

[SZ84] Y.S. Sherif and S.S. Zahir. On power-line carrier communication (PLC). *Microelectronics and Reliability*, 24(4):781–791, 1984.

[TXS+04] O.K. Tonguz, A.E. Xhafa, D.D. Stancil, A.G. Cepni, P.V. Nikitin, and D. Brodtkorb. A simple path-loss prediction model for HVAC systems. *IEEE Transactions on Vehicular Technology*, 53(4):1203–1214, July 2004.

[WCH88] D.J. Wilcox, M. Conlon, and W.G. Hurley. Calculation of self and mutual impedances for coils on ferromagnetic cores. *Physical Science, Measurement and Instrumentation, Management and Education - Reviews, IEE Proceedings A*, 135(7):470–476, September 1988.

[WHD+08] Ingolf Wasserman, Detlef Hahn, Hai Nguyen Dang, Hanno Reckmann, and John Macpherson. Mud-pulse telemetry sees step-change improvement with oscillating shear valves. *Oil & Gas Journal*, 106(24):39–40, June 2008.

[wik10] Wikipedia: Electrical conductivity, August 2010. Available from: `http://en.wikipedia.org/wiki/Electrical\_conductivity`.

[Wil02] Andreas Willig. Analysis of the PROFIBUS token passing protocol over wireless links. In *IEEE International Symposium on Industrial Electronics*, volume 1, pages 56–60, 2002.

[Wil03] Andreas Willig. Polling-based MAC protocols for improving real-time performance in a wireless PROFIBUS. *IEEE Transactions on Industrial Electronics*, 50(4):100–112, August 2003.

[Wil08] Andreas Willig. Recent and emerging topics in wireless industrial communications: A selection. *IEEE Transactions on Industrial Informatics*, 4(2):102–124, May 2008.

[WL19] J.A. Willoughby and P.D. Lowell. Development of loop aerial for submarine radio communication. *American Physics Society Journal*, 14(2):193, 1919.

[WMP09] Thomas Watteyne, Ankur Mehta, and Kris Pister. Reliability through frequency diversity: why channel hopping makes sense. In *Proceedings of the 6th ACM symposium on Performance evaluation of wireless ad hoc, sensor, and ubiquitous networks*, 2009.

[WMRD11] Thomas Watteyne, Antonella Molinaro, Maria Grazia Richichi, and Mischa Dohler. From MANET to IETF ROLL standardization: A paradigm shift in WSN routing protocols. *IEEE Communications Surveys & Tutorials*, 13(4):688–707, Fourth Quarter 2011.

[WMW05] Andreas Willig, Kirsten Matheus, and Adam Wolisz. Wireless technology in industrial networks. *Proceedings of the IEEE*, 93(6):1130–1151, May 2005.

[WSL10] Ludwig Winkel, Zafer Sahinoglu, and Liang Li. Ieee std 802.15.4e-d0.01, March 2010.

[WW01] Andreas Willig and Adam Wolisz. Ring stability of the PROFIBUS token-passing protocol over error-prone links. *IEEE Transactions on Industrial Electronics*, 48(5):1025–1033, October 2001.

[YF04] Ossama Younis and Sonia Fahmy. Heed: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. *IEEE Transactions on Mobile Computing*, 3(4):366–379, October-December 2004.

Contributions

Paper A

Magneto-Inductive Communication among Pumps in a District Heating System

Soroush Afkhami Meybodi, Pablo Pardo, and Mischa Dohler

This paper is published in:
The 9th International Symposium on Antennas, Propagation, and EM Theory
December 2010, Guangzhou, China

Abstract

Realizing distributed control for a large scale district heating plant needs a metropolitan wireless communication framework among its pumps. The main challenge at PHY layer here is coverage rather than interference. This paper reviews the potential methods and proposes the use of magnetic induction via underground pipes as very long ferromagnetic cores. Feasibility of this method is verified by simulations.

1 Introduction

Some below ground urban utilities, like water distribution [1, 2, 3] and district heating [4] are among the intriguing applications for joint control and data communication systems. It is envisioned that a co-design could result in considerable resource savings by facilitating distributed automatic control of these geographically dispersed plants. Our focus in this paper is on propagation aspects of the communication framework for district heating system case study.

District heating is a way of providing heat demands to closely located buildings by supplying hot water to them and cycling the return cold water into central heating plants via dedicated pair-wise pipelines which form a closed hydraulic system. The hydraulic network is driven by centrifugal pumps. Pipes are buried underground and pumps are usually placed in the basement of buildings along the pipes and are wired to local pressure sensors to provide local control of the hydraulic dynamics. In the newly proposed piping layout for district heating systems [5] which is designed in sake of less heat loss, cooperation among the pumps is needed to maintain the stability of the hydraulic network. This urges data communication among pumps.

The main feasibility issues with the available data communication methods for underground pumps are 1) wiring costs for wired communication systems since the piping infrastructure is already installed; 2) dependency on a third party infrastructure for cellular and power line communications plus lack of quality of services for closed loop control. There are cellular solutions for monitoring purpose available, but closed loop control needs more frequent messaging which increases the price and introduces real-time problems; and 3) very limited trans-mission range for electromagnetic (EM) wireless solutions.

Regarding EM solutions, the maximum transmission range depends on transmit power, antenna gains, propagation losses, receiver sensitivity, and losses in the transmitter and receiver circuits. Propagation losses play the dominant role in underground applications. They consist of path loss, shadowing and fading. Pathloss highly depends on dielectric properties of the material and is significantly higher in soil than that of free space. For instance, if EM waves are employed for through-the-soil single path communication, according to [6], a maximum range of 3.9 m is expected for an un-coded bit error rate (BER) of 10^{-3} considering 30 dBm transmit power and PSK modulation with no shadowing and fading. This range is increased by about 1 m if the burial depth is not that much and the multipath effect could be used constructively [6].

Similar results for the maximum achievable transmission range have been reported in channel model studies [7]. For instance, indoor non-line-of-sight (NLOS) scenarios suggest a 5 dB pathloss for signal penetration through thin walls and 12 dB for thick walls.

Although thickness of the walls are not explicitly mentioned, the results are comparable with [6], assuming a typical 100 dB tolerable propagation loss per link.

An alternative idea could be to use the infrastructure itself, i.e. underground pipes, as EM waveguides to alleviate path loss similar to what was accomplished in [8] for air-filled ducts of heating, ventilation and air conditioning (HVAC) systems. However, high conductivity of water in pipes compared to air in ducts, which is in relative order of 2×10^{11} S/m [9, 10], results in high absorption of EM waves in water and anticipates no good coverage range. More clearly, EM propagation in soil is much better than in water.

Magnetic Induction (MI) is another propagation technique, especially useful for through-the-soil communication [6, 11, 12]. Unlike EM, there is no tangible difference between MI propagation in air, water and soil because of the similar magnetic permeability values for these environments. Nevertheless, MI attenuates with $1/r^6$ in a magnetically homogenous environment, where r is the distance between the transmitter and receiver. Therefore, it is mostly suitable for short range applications.

A completely different solution could be to use water itself as the communication medium by employing acoustic waves [13, 14]. An interesting issue is the absence of inter-system interference simply because there is no other source of data transmission in that proprietary medium. Moreover, it would be possible to extract other information e.g. leakage detection in pipelines by monitoring unexpected permanent noise which is an aftermath of leakage of water through a pipe breakage.

The most serious drawback to use acoustic method to communicate among pumps is the very high price of the commercially available hydrophones that is even comparable to the price of the pump itself. This is in contrast with the low prices of wireless motes that are being used ubiquitously in wireless sensor networks. Moreover, the acoustic noise in the channel is at its maximum level when we need a high control bandwidth, i.e. more acoustic noise when pressure changes in the pipes. Therefore, the communication links have the least reliability when their maximum capacity is needed.

In summary, to deal with this specific application, we tried the use of pipes as the communication medium and MI as the key technology. District heating steel pipes act as extended ferromagnetic cores between neighboring pumps. This makes it possible to considerably increase the coverage range compared to the result reported in [6, 11]. Mutual inductances and power transmission ratio are found by simulations for different materials with a wide range of magnetic permeability. The results are derived for both single pipe, e.g. water distribution systems, and double pipe, i.e. district heating.

The rest of the paper is as follows. After clarifying coverage range requirements in Section 2 and introducing magneto-inductive communication in Section 3, we present the simulation results in Section 4. Section 5 concludes the paper.

2 Problem Statement

Figure 6.1 shows a typical urban area with district heating pipes. Each line represents a pair of pipes, i.e. both the forward and return pipes. Pumps are shown as small red dots on the pipes. They are placed in basement of buildings on the forward pipe in depth of one to two meters underground. Basements can be thought of as underground air-filled chambers with concrete walls.

Connecting pipes are buried about one meter below the surface along roads in sand

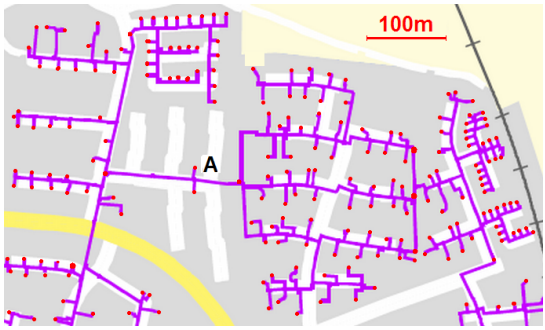

Figure 6.1: Pipes and pumps in an urban area

filled channels. Pipes are made of ferromagnetic steel. They are all welded together and thermally insulated from the environment with polyester jackets filled with a thermal insulator [9].

As a crucial step, distances among pumps should be studied. To this end, physical location of 180 pumps in a sample area was analyzed. Figure 6.2 shows the cumulative distribution function (CDF) of distances for different number of nearest neighbors from one to five.

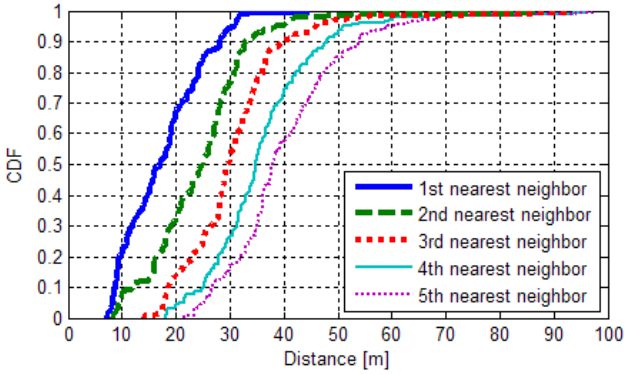

Figure 6.2: CDF of the first five nearest neighbor nodes

For instance, it shows that the distance for 50% of the pumps from their respective nearest neighbor is about 17 m. This increases to 25 m for the second nearest neighbor and so on.

It is worth saying that if all the 1^{st} (i^{th}) nearest neighbors are well in range, it does not guarantee the full connectivity of the network because there may exist isolated sub-sections made of (i+1) nodes. Deriving a connectivity measure is not in the scope of this paper and hence we have investigated the feasibility of communication for the farthest possible distance which is considered equivalent to a signal loss of 100 dB, the dynamic range of typically used embedded radios.

3 Magnetic Induction with Extended Ferromagnetic Cores

MI is a well known technique for short range communications. There are patents and commercial products on the market that employ this technology. Its usefulness for underground applications has been exposed in [6, 11, 12].

Two wire coils are used in MI as the transmitter and receiver. Electric current in the transmitter coil creates a magnetic field. When it changes, it will produce an electric current in the receiver coil due to mutual inductance between the two coils. Therefore, finding mutual inductance is the key point to derive quantitative results.

Two different approaches to finding the mutual inductance are: 1) using approximate closed form expression [11, 15] and 2) exact closed integrations suitable for implementation in numerical solvers [16]. In this paper, we have used a finite element solver to numerically find self and mutual inductances for coils with ferromagnetic cores. However, the following formulas for air-core coils are borrowed from [11] to compare the improvement resulted by using a ferromagnetic core.

$$L_t \approx \frac{1}{2}\mu\pi N_t^2 a_t \qquad L_r \approx \frac{1}{2}\mu\pi N_r^2 a_r \tag{6.1}$$

$$M \approx \mu\pi N_t N_r \frac{a_t^2 a_r^2}{2r^3} \tag{6.2}$$

L_t and L_r are self inductance of the transmitter and receiver coils. M is the mutual inductance. μ is the permeability of the homogenous surrounding, N_t and N_r the number of turns of coils, a_t and a_r the radii of coils and r is the distance between the two coils.

Self and mutual inductances are used to find the ratio between the input complex power to the transmitter coil (P_t) and the output complex power into an impedance load (P_r) in order to evaluate MI pathloss. Using the equivalent electric circuit of coupled inductors, and matching the load impedance to receive the highest power, the exact expression for power transmission ratio is derived in [11] as:

$$\frac{P_r}{P_t} = \frac{\left[j\omega M/(R_t + j\omega L_t)\right]^2}{4R_r + 4R_t\omega^2 M^2/(R_t^2 + \omega^2 L_t^2)} \cdot \left(R_t + j\omega L_t + \frac{\omega^2 M^2}{2R_r + \omega^2 M^2/(R_t - j\omega L_t)}\right) \tag{6.3}$$

where ω is the angular velocity of the signal, and R_t and R_r are the electrical resistance of coils. Evaluating (6.3) is the final step to assess the feasibility of MI communication.

4 Simulation Results

The CST EM Studio®magneto-static solver is used for finding self and mutual inductances. Figure 6.3 shows two different models, suitable for single pipe and double pipe applications, respectively.

In contrast to the real practice, straight open ended pipes are considered in the simulations. To minimize the effect of this mismatch on the magnetic field between coils, the coils should not be too close to the ends of the pipe. For example, if we are to find the mutual inductance with a specific accuracy, the pipe should be long enough to the extent that repeating the simulation with a longer pipe does not affect the result. We have found

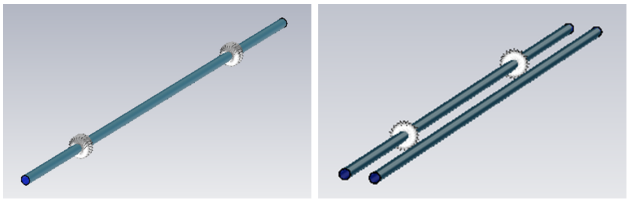

Figure 6.3: Single pipe and double pipe structures

it by trial and error. Other parameters of the pipe are: 1) outer diameter 38 mm, and 2) thickness 3.65 mm, conforming to DIN 2440 standard. Pipes are filled with water.

Figure 6.4 shows how the self inductance of coils depends on the magnetic permeability of the pipes varying from μ_0 to $750\mu_0$. L_t and L_r are alike because of the symmetry of the structures. N_t and N_r are equal to 100. As it is plausible, the distance between two coils has negligible effect on the self inductance. Logarithmic scale on Y axis is used just for better visualization.

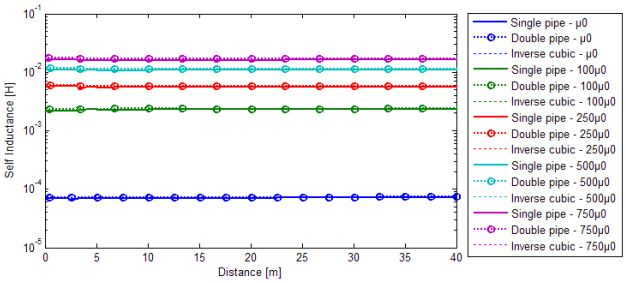

Figure 6.4: Self inductance L of coils

On the other hand, mutual inductance is expected to decrease approximately by a factor of $1/r^3$ (6.2) in magnetically homogenous surroundings. Figure 6.5 compares this hypothesis with the simulation results. It indicates that the mutual inductance of air-core coils reduces faster for small distances after starting from the same point at zero distance. At the same time, it shows that for larger distances, the rates of diminution are similar irrespective of the core material.

References [6] and [11] predict a range no better than 3 m for 100 dB admissible pathloss, but Figure 6.5 promises enhanced ranges by showing much higher mutual inductance compared to air-core coils. It is also shown that there is no tangible difference between single pipe and double pipe structures.

The next step is to evaluate (6.3) and calculate pathloss between the transmitter and receiver coils. Figure 6.6 shows the pathloss, for operating frequency 3000 Hz. This frequency is not licensed for radio communications and is low enough to ignore electrical field [12]. It is assumed that the wire unit length resistance is 0.34 ω/m, conforming to AWG 30 standard.

The dotted line represents the air-core case, where having a viable communication link of approximately 2.5 m is expected. The coverage range is considerably increased

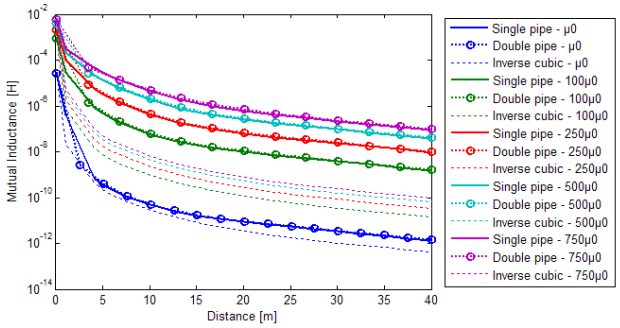

Figure 6.5: Comparing the diminution of mutual inductance M with $1/r^3$

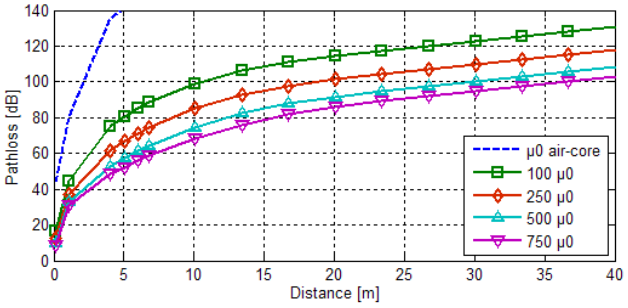

Figure 6.6: Pathloss versus distance between coils at 3000 Hz

by using a ferromagnetic core. This suggests that magneto-inductive communication may be used for urban utilities like water distribution and district heating where iron and steel pipes are used to build the pipelines.

Last but not least, when comparing Figure 6.6 with the reported results in [6] and [11], one must consider the large difference between operating frequencies. We have used a much lower frequency which results in a higher pathloss based on (6.3). However, it reduces core losses and far filed effects at the same time.

5 Concluding Discussion

Magnetic permeability of ferromagnetic steel pipes varies in a wide range up to several hundred μ_0 depending on their metallurgic composition. This gives a maximum coverage range from approximately 20 to 40 m for $250\mu_0$ to $750\mu_0$, provided that a maximum 100 dB of pathloss is tolerable. Back to the results in Section II and taking $500\mu_0$ into account, more than 95% of the pumps have at least one neighbor at reach. This value decreases to 75% and 50%, for the second and third nearest neighbors.

The simulation results are promising, but in practice, pipes have bents, branches and fittings. Therefore, our future work is to design and run experiments to verify these achieved results in real world where we also deal with extra factors like eddy current losses, B-H nonlinearity and B-H frequency dependent hysteresis losses. Although the operating frequency is quite small, i.e. 3 kHz, these losses could still be considerable.

References

[1] WIDE: Decentralized and Wireless Control of Large-Scale Systems, WP2 & WP5. [Online]. Available: http://ist-wide.dii.unisi.it

[2] Neptune project, WP1.1. [Online]. Available: www.shef.ac.uk/neptune

[3] WINES: Wireless Intelligent Networked Systems. [Online]. Available: http://www-civ.eng.cam.ac.uk/geotech_new/WinesInfrastructure/

[4] P^3C: Plug and Play Process Control, WP1. [Online]. Available: http://vbn.aau.dk/en/projects/plug-and-play-process-control-p3c(cdd30235-5d31-4f94-bd17-c29028934471).html

[5] B. Bohm, N. K. Vejen, J. Rasmussen, N. Bidstrup, K. P. Christensen, F. Bruus, and H. Kristjansson, "EFP-2001 Fjernvarmeforsyning af lavenergiomraader," Energistyrelsen, Tech. Rep., March 2004, in Danish.

[6] I. Akyildiz, Z. Sun, and M. Vuran, "Signal propagation techniques for wireless underground communication networks," *Physical Communication*, vol. 2, no. 3, pp. 167–183, September 2009.

[7] P. Kyosti, J. Meinila, L. Hentila, X. Zhao, T. Jamsa, C. Schneider, M. Narandzic, M. Milojevic, A. Hong, J. Ylitalo, V. Holappa, M. Alatossava, R. Bultitude, Y. de Jong, and T. Rautiainen. (2008, 2) WINNER II channel models, part I, D1.1.2 V1.2. [Online]. Available: http://www.ist-winner.org/WINNER2-Deliverables/D1.1.2.zip

[8] O. Tonguz, A. Xhafa, D. Stancil, A. Cepni, P. Nikitin, and D. Brodtkorb, "A simple path-loss prediction model for HVAC systems," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 4, pp. 1203–1214, July 2004.

[9] Danfoss heating book - 8 steps to control of heating systems. [Online]. Available: http://heating.danfoss.com/Content/61078BB5-1E17-4CA2-AC49-8A7CAC2BA687_MNU17424997_SIT54.html

[10] (2010, August) Wikipedia: Electrical conductivity. [Online]. Available: http://en.wikipedia.org/wiki/Electrical_conductivity

[11] Z. Sun and I. Akyildiz, "Underground wireless communication using magnetic induction," in *In Proc. IEEE ICC*, Dresden, Germany, June 2009.

[12] J. Sojdehei, P. Wrathall, and D. Dinn, "Magneto-inductive (MI) communications," in *Oceans, MTS/IEEE Conference and Exhibition*, Honolulu, HI, USA, November 2001, pp. 513–519.

[13] G. Kokossalakis, "Acoustic data communication system for in-pipe wireless sensor networks," Ph.D. dissertation, Massachusetts Institute of Technology, February 2006.

[14] R. Long, P. Cawley, and M. Lowe, "Acoustic wave propagation in buried iron water pipes," in *Proceedings A: Mathematical, Physical & Engineering Sciences*, vol. 459, no. 2039. The Royal Society, November 2003, pp. 2749–2770.

[15] J. Bhagwan and F. Trofimenkoff, "Mutual inductance of coils on a cylinder of infinite length," *IEEE Transactions on Magnetics*, vol. 16, no. 2, pp. 477–479, March 1980.

[16] D. Wilcox, M. Conlon, and W. Hurley, "Calculation of self and mutual impedances for coils on ferromagnetic cores," *Physical Science, Measurement and Instrumentation, Management and Education - Reviews, IEE Proceedings A*, vol. 135, no. 7, pp. 470–476, September 1988.

Paper B

Magneto-Inductive Underground Communications in a District Heating System

Soroush Afkhami Meybodi, Jens Dalsgaard Nielsen, Jan Dimon Bendtsen, and Mischa Dohler

This paper is published in: Proceedings of the
IEEE International Conference on Communications, ICC2011
June 2011, Kyoto, Japan

Abstract

Feasibility of underground data communications is investigated by employing magnetic induction as the key technology at physical layer. Realizing an underground wireless sensor network for a district heating plant motivates this research problem. The main contribution of the paper is to find the optimal design for transmitter and receiver coil antennas, assuming pre-defined wire gauge and length at both sides. Analytical results show that the transmitter coil should have the smallest possible radius while the converse is true for the receiver coil.

1 Introduction

Wireless communications in underground and confined areas have always been a challenging problem due to the limited coverage range. We faced this problem on our way to realize a distributed control system for district heating networks. The goal is to create reliable communication links among pumps of an urban-wide hydraulic network. Pumps are located approximately 2 m underground in buildings' basements and are physically connected together via welded steel pipes. A previous study [1] has shown that a 40 m coverage range is fairly acceptable. It provides at least one reachable neighbor for each pump, while 60% of the pumps have five accessible neighbors in a typical piping layout.

Several technologies have been examined by researchers from various fields for underground and even in-pipe communications. A thorough survey is presented in [2] where the following technologies are investigated in details.

- Radio Frequency (RF) Electromagnetic (EM) waves through-the-soil
- RF EM waves for combined underground and above-ground scenarios
- Pipelines as EM waveguides
- Acoustic waves through the pipelines
- Acoustic waves through the district heating water
- Power Line Communications
- Employing electrical conductivity of the steel pipes
- Incorporating Cell Phone infrastructure
- Magnetic Induction (MI)

From the above choices, all but the last one, turned out to have serious shortcomings [2]. Therefore, the focus of this paper is on MI as the potential winning candidate.

MI, to be described in Section II, has already found several applications in practice. It is used in very low power wireless microphones and headsets that are used by security forces [3] in frequency ranges 11-15 MHz and distances up to 3 m providing a high level of security by involving quasi-static fields, to be described later. Lower frequencies of 300 Hz to 4 kHz are employed for distances up to a few kilometers in soil or water medium. They have been used in underground and underwater voice and data communications for mining and military purposes, underwater navigation, remote triggering of

underground and underwater mines and ammunition, seismic fence, and submarine communication systems [4, 5]. We will see in Section II, that in such cases the communication media are hardly penetrable by higher frequencies.

The main contribution of this paper is presented in Section III, where we derive the optimal design criteria for transmitter and receiver antennas of a magneto-inductive communication system by using computer simulations which are based on analytical formulas. Section IV concludes the paper by giving some calculations regarding applicability of the MI technology in our original district heating problem.

2 Magnetic induction

2.1 MI vs. RF EM waves

MI has been around since 1920's being used as a means of communication between submarines [5]. There is no fundamental difference between MI and RF EM waves. MI also employs alternating Electric ($\mathbf{E}$) and Magnetic ($\mathbf{B}$) fields as the data carrier, but it uses low frequencies and low power such that the communication range falls off within the distance predominated by the *static* or *quasi-static* fields, as described in the sequel.

$\mathbf{E}$ and $\mathbf{B}$ fields which are created by an alternating electrical current in an antenna can be estimated by the summation of several terms that have different dependencies to the distance *r* from the antenna, e.g. $1/r^3$, $1/r^2$, and $1/r$. In the vicinity of the antenna, fields are very complex (to be shown in the simulations in Section III). Gradually, they are predominated by $1/r^3$ terms. Up to this zone, is called the *static field* domain. Other terms predominate at larger distances and are called *inductance field* (*quasi-static field*) and *radiation field*, respectively [6]. The latter accounts for EM radiation.

To define which field component is the leading one, r should be compared with $k = \omega/v$ where ω is the angular frequency of $\mathbf{E}$ and $\mathbf{B}$ fields, v is the speed of EM waves in the medium, and k is called the *wave number*. $k.r \gg 1$ and $k.r \ll 1$ imply that the radiation and static fields are dominant, respectively. Otherwise, the field is quasi-static.

Static and quasi-static fields decay much faster than the radiation field with increasing distance from the current source. This property of non-radiating fields have made them intriguing for military applications where security is a concern, but it is of little value in our district heating problem. Actually, it is only the choice of the operating frequency which leaves us in the non-radiating field domain.

Why should we choose such low operating frequencies? The answer lies in the dielectric properties of lossy materials like wet soil. When materials are influenced by alternating $\mathbf{E}$ and $\mathbf{B}$ fields, their electric susceptibility (χ_e) can be stated as a complex function of frequency:

$$\chi_e(\omega) = \chi'_e(\omega) - j\chi''_e(\omega) \tag{7.1}$$

The general behavior of the real part $\chi'_e(\omega)$ and the imaginary part $\chi''_e(\omega)$ for most of the materials is sketched in Fig. 7.1 [7]. $\chi'_e(\omega)$ increases most of the times with respect to ω, except for short intervals where it drastically decreases. These intervals coincide with local maxima in $\chi''_e(\omega)$. $\chi'_e(\omega)$ defines the velocity of $\mathbf{E}$ and $\mathbf{B}$ fields in the matter, while $\chi''_e(\omega)$ accounts for attenuation due to molecular resonance phenomena.

It provides an extra *alternating field conductivity* (σ_a) besides the common *static field conductivity* (σ_s) at resonance frequencies. The equivalent conductivity (σ_e) can be stated

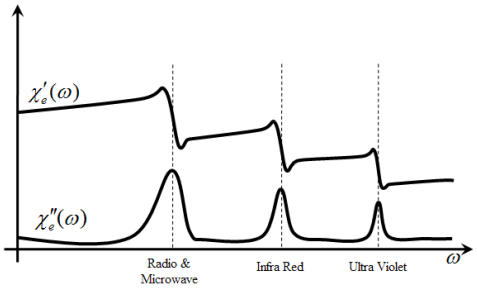

Figure 7.1: Electrical susceptibility of matters versus frequency

as:

$$\sigma_e = \sigma_s + \sigma_a = \sigma_s + \omega(\epsilon_0 \chi_e''(\omega)) \tag{7.2}$$

Equation (7.2) and Fig. 7.1 imply that, $\mathbf{E}$ and $\mathbf{B}$ fields should have low frequencies to alleviate the total conductivity, and hence the attenuation. The lower frequency, the less attenuation.

A starting point to choose the appropriate operating frequency is to compare our problem with other MI applications. Knowing that the *index of refraction* (n) of EM waves in soil is $\sim$1.5 [8], the speed of EM waves in soil is given as $v = c/n \approx 2\times10^8$ m/sec. Therefore, for distances of 10 m to 70 m and working frequencies of 2 kHz to 9 kHz, we have $k.r \in$[0.006,0.01] which falls well in the non-radiating field domains and does not interfere with any licensed frequencies.

The above discussion explains why MI is a potential solution to our problem. Furthermore, in non-radiating fields, there holds no relation such as $|\mathbf{E}|/|\mathbf{B}| = v$. Thus, $\mathbf{B}$ can be substantially larger than $\mathbf{E}$. Therefore, in a MI system, multipath effects is minimized due to fairly similar values of *magnetic permeability* (μ) for most of the materials.

2.2 System Components

Fig. 7.2 shows the basic elements of an MI communication system. Coil antennas are used at the transmitter and receiver sides. Coils have low radiation resistance and transmit very little real power. Therefore, they have low performance for radiation field. Instead, a coil antenna has a high reactive power going back and forth to the antenna and the surrounding space in each cycle. This property is desirable for MI, in contrast with typical RF systems where reactive power is kept as low as possible.

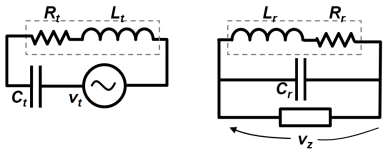

Figure 7.2: MI transmitter and receiver circuits

A coil antenna operating at low frequencies can be well modeled as an ideal inductance in series with a resistance which is equal to the resistance of the coil wire. We assume that the operating frequency is low enough, hence ignore the self capacitance across the coil in its lumped model.

At the transmitter, our objective is to produce the strongest possible $\mathbf{B}$ field at a considerable distance from the antenna. This will be achieved when the electrical current in the coil (I_t) is maximized. Therefore, a capacitor C_t should be connected to the coil in series and bring it to resonance such that $I_t = v_t/R_t$. In practice, a voltage source with low output impedance should be used. For example, an audio amplifier in conjunction with a standard signal generator works well [2].

The receiver coil, on the other hand, should be sensitive to the slightest change of the passing magnetic flux and produce the highest voltage. The induced back-emf v_r can be calculated in a multi-turn coil by:

$$v_r = \omega . \sum_{i=1}^{N_r} \left[\int_{A_i} \bar{\mathbf{B}} . \hat{\mathbf{n}} d(s) \right] \tag{7.3}$$

where ω is the angular frequency of I_t, A_i the area of individual turns of the Rx coil, N_r the number of turns of the Rx coil, and $\bar{\mathbf{B}}$ is the RMS magnetic field passing through it. If $\bar{\mathbf{B}}$ has a constant amplitude at all points on the Rx coil surface and is parallel to $\hat{\mathbf{n}}$, which has the unit length and is perpendicular to the coil surface, (7.3) will be simplified as follows:

$$v_r = \omega . |\bar{\mathbf{B}}| . \sum_{i=1}^{N_r} A_i \tag{7.4}$$

which shows that $\sum_{i=1}^{N_r} A_i$ should be maximized in order to get the highest possible back-emf. Furthermore, if a capacitor C_r is connected to both ends of the coil such that it brings the coil into resonance (see Fig. 7.2), the followings hold:

$$\begin{aligned} C_r &= \frac{1}{\omega^2 L_r} \\ v_z &\approx \frac{1}{jC_r\omega} I_r = -j\omega L_r (\frac{v_r}{R_r}) = -jQv_r \end{aligned} \tag{7.5}$$

provided that the input impedance of the receiver amplifier/filter is very large and the *quality factor* (Q) of the coil is defined as $Q = \omega L_r/R_r$. Therefore, if the Rx coil has a high Q, the passive band-pass receiver circuit acts as a pre-amplifier.

Combining (7.4) and (7.5), the design objective in the Rx antenna is to maximize the following expression:

$$\frac{L_r}{R_r} . \sum_{i=1}^{N_r} A_i \tag{7.6}$$

As another approach, coupling between the transmitter and receiver antennas can be formulated as mutual inductance between the two coils [9]. In district heating and water distribution systems, it is possible to amplify the mutual inductance by employing pipes as ferromagnetic cores of the transceiver antennas. This will increase the self and mutual

inductance in expense of adding core losses. We have shown in [1] that the mutual inductance can be enhanced by several orders of magnitude when coils are wound around steel pipes. However, iron losses counteract mutual inductance enhancement and we have used air core coils throughout this paper.

Employing intermediate coils as magneto-inductive waveguides is another technique to increase the range in specific directions and are investigated in [10], and [11]. However, it is not practical in our problem to insert intermediate coils because of the lack of pipes accessibility at regular distance intervals.

3 Coil Antenna Design

Given a specific wire gauge R, and length l, an interesting problem is to find how to wind the coils such that the communication range is extended as far as possible. Assumptions on wire gauge and length are reasonable since they define the cost and weight of the coils. They hold throughout this section.

We start by introducing the analytical expression for the $\mathbf{B}$ field in all points of the space, created by a single turn current loop as shown in Fig. 7.3. The formula is derived by calculating the magnetic vector potential first, and taking the curl of it to find $\mathbf{B}$ [12].

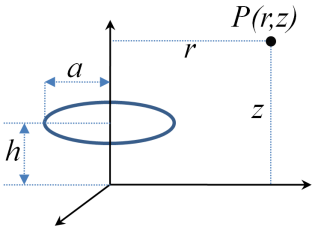

Figure 7.3: A current loop

$$\mathbf{B}(r,z) = \frac{\mu I k}{4\pi\sqrt{ar^3}}\left[-(z-h)\left(K(k) - \frac{2-k^2}{2(1-k^2)}E(k)\right)\hat{\mathbf{r}} + r\left(K(k) + \frac{k^2(r+a)-2r}{2r(1-k^2)}E(k)\right)\hat{\mathbf{z}}\right] \quad (7.7)$$

$\mathbf{B}$ is expressed in terms of complete elliptic integral functions of first and second kind, i.e. $K(k)$ and $E(k)$, where their argument k is defined as:

$$k = \sqrt{\frac{4ar}{(r+a)^2 + (z-h)^2}}$$

All other parameters of (7.7) are expressed in Fig. 7.3. If $\mathbf{B}$ is only sought on z-axis, (7.7) will be significantly simplified. In that case, it would be more convenient to directly use the Biot-Savart law to calculate on-axis $\mathbf{B}$ field.

3.1 Relative positioning of Tx and Rx antennas

The graphical representation of $|\mathbf{B}|$ for a multi-turn coil is given in Fig. 7.4 as a contour plot at two zoom levels. The coil is not shown in this plot, but it is laid in $z = 0$ plane and centered at the origin, with $a = 0.1$ m.

The field is calculated at each point in space by summing up (7.7) for individual wire turns, considering the exact radius and location for each turn of wire.

It is shown that $|\mathbf{B}|$ is larger around the coil and on the coil's axis. An interesting behavior is observed on $z = 0$ plane where $|\mathbf{B}|$ has a local minimum at a radius of approximately 2.4 a. This is due to the $\mathbf{B}$ field generated by the currents flowing in farther parts of the loop which counteract the $\mathbf{B}$ field generated by the currents flowing in the closer parts. At larger distances from the coil, $|\mathbf{B}|$ starts to degrade monotonically.

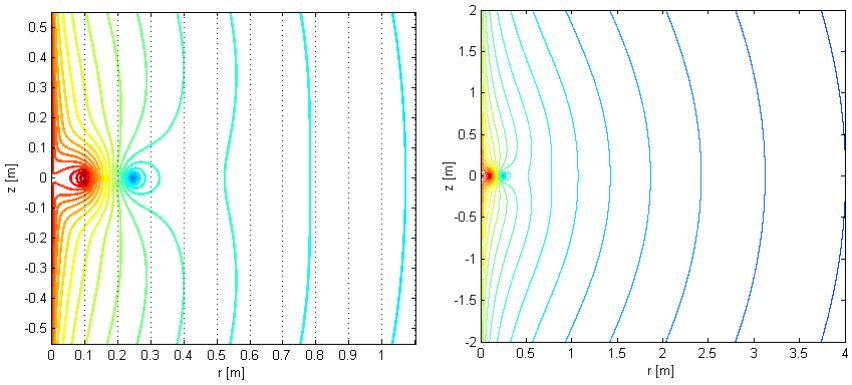

Figure 7.4: Contour plots of $|\mathbf{B}|$ in the space surrounding the Tx antenna at two zoom levels

Since the contours in Fig. 7.4 (right) do not follow an exact spherical pattern, there are variations in $|\mathbf{B}|$ at fixed distances from the Tx antenna. It is clear that $|\mathbf{B}|$ is larger around the z-axis. This behavior is quantified in Fig. 7.5, in which $|\mathbf{B}|$ is drawn at three different fixed distances of 1 m, 10 m, and 100 m from the Tx antenna with $a = 0.1$ m. Each curve starts from vicinity of the z-axis, associated with $\theta = 0.1^\circ$ to $z = 0$ plane, associated with $\theta = 90^\circ$, where θ is illustrated in Fig. 7.5. The results are shown in log-log scale in sake of highlighting the differences. Moreover, the curves are shifted vertically such that they all coincide at $\theta = 90^\circ$. Fig. 7.5 shows that $|\mathbf{B}|$ can vary up to approximately 20 dB at the Rx antenna based on positioning of the Tx antenna if $\theta \in [1^\circ, 90^\circ]$. It also shows that $|\mathbf{B}|$ is more sensitive for small values of θ.

Note that Fig. 7.5 does not say anything about the angles between Tx and Rx coils' axes, which is important in calculating the magnetic flux passing through the Rx coil.

3.2 Cross section shapes

Given the mean radius of the coil a, and hence the number of its turns N, ($N = l/(2\pi a)$), we have compared rectangular cross section shapes with different side proportions to see its effect on the Tx and Rx antennas. These are ranging from a single layered coil in

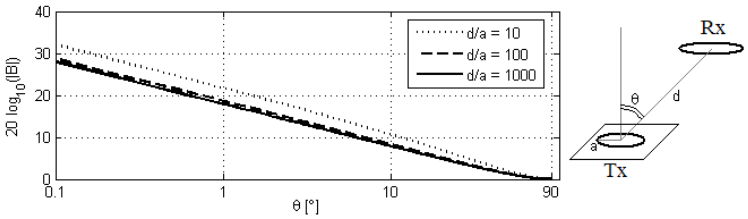

Figure 7.5: Normalized $|\mathbf{B}|$ vs. Rx position with respect to the Tx axis

which $b = 2R$ and $c = Rl/(\pi a)$ to a coil with the maximum number of winding layers which is made when $b \leq 2a$ and $c \geq R^2 l/(\pi a^2)$. The latter holds, provided that each wire occupies a square surface of area $4R^2$ at the cross section of the coil. The most compact wiring, results in hexagonal surface assignment to each wire turn and $c \geq \frac{\sqrt{3}R^2 l}{2\pi a^2}$. All of the above inequalities convert to equalities if the wire is very long such that no hole remains in the center of the multi-turn coil.

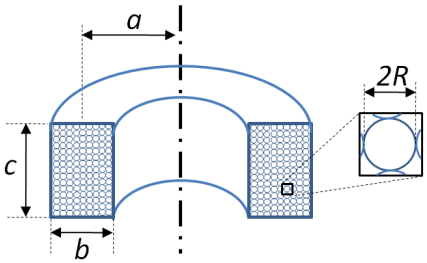

Figure 7.6: Dimensions of a multilayer coil with rectangular cross section

At the Tx coil, the difference between $|\mathbf{B}|$ of a single layer and a square coil at $z = 0$ plane is in order of 0.01% of the $|\mathbf{B}|$ created by either of them. In all other points in the space, this trifling difference is even more negligible. Therefore, the cross section shape of the Tx antenna is not tangibly influential on $|\mathbf{B}|$.

At the Rx coil, on the other hand, the coil's cross section shape is quite effective mostly due to its effect on the self inductance of the coil (L_r). This is explicitly presented in (7.6). Since the length and the mean radius of the coil are pre-defined, $\sum_{i=1}^{N_r} A_i$ is constant and L_r must be maximized.

Maximization of self inductance of a coil was first posed by Maxwell in [13]. He has shown that the *geometric mean distance* between the conductors of an inductor should be minimized in order to get the maximum inductance. This is achieved when the cross section is circular. For rectangular shapes which are easier to manufacture, the best shape is a square, i.e. $b = c$ (See Fig. 7.6). This result is verified by simulations in Fig. 7.7.

The coil in this simulation is made up with 150 m of wire with gauge $R = 0.3$ mm. The cross section is rectangular and the mean radius is $a = 0.20$ m. Consequently, the number of turns N_r turns out to be 119 which yields 11 winding layers for a square cross section. Fig. 7.7 confirms this fact by showing a peak at N_r =11. It also shows that a single layer Rx coil has only 60% performance of the square shaped Rx coil.

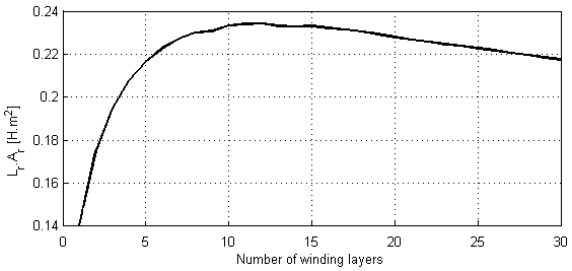

Figure 7.7: emf variations in the Rx coil as a function of cross section shape

In the above simulation, L_r is calculated by injecting a current I into the Rx coil and computing the induced magnetic flux Φ_i in each turn due to the current in all of the turns. Then, $L_i.I = \Phi_i$ gives L_i and $L_r = \sum_{i=1}^{N_r} L_i$. Note that, in calculating the total self inductance of the Rx coil, we only added up L_i of individual turns which convey the self inductance of that turn plus its mutual inductance to all other turns.

3.3 Mean radius vs. Number of turns

Since the total wire length ($l = 2\pi aN$) is pre-defined, aN is constant and we should make a trade-off between a and N. At the Tx antenna, Fig. 7.8 exhibits the variations in $|\mathbf{B}|$ at $z = 0$ as r increases from 1 m up to 50 m, for different values of a from 0.03 m to 0.25 m and square cross section.The lowest value is associated with the physical limitation of $a > b/2$ (see Fig. 7.6 for clarification).

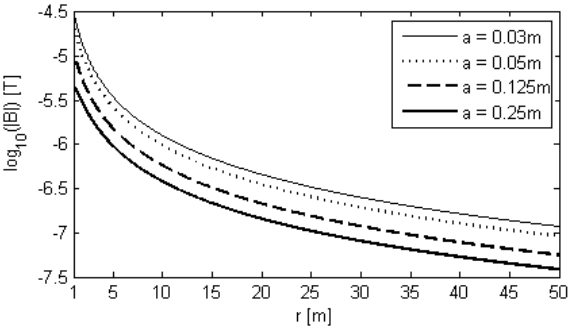

Figure 7.8: Variations in $|\mathbf{B}|$ as r increases

Fig. 7.8 suggests that a should be as small as possible to have the maximum $|\mathbf{B}|$ at large distances. The same results were achieved when $z \neq 0$ and also on any straight line with slope α such that $z = \alpha.r$. No matter how large a finite α is, the above hypothesis holds true and the smallest mean radius gives the maximum $|\mathbf{B}|$ at distances of interest. The only exception is $\alpha = \infty$, i.e. $|\mathbf{B}|$ on the z axis in which the converse is true, but it does not have any practical importance.

For the Rx coil, (7.6) is depicted in Fig. 7.9 as a function of a. For all values of a, a square cross section is considered.

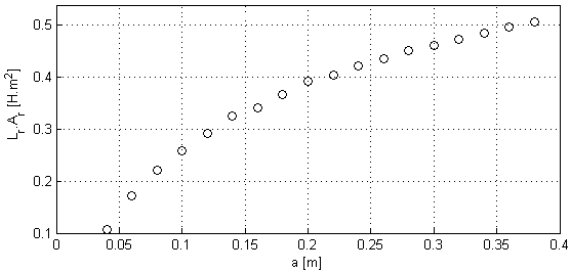

Figure 7.9: Variation in v_z at the receiver coil as a increases

Fig. 7.9 admits that for the Rx coil, larger a produces higher voltages, no matter if the number of turns or L_r is decreased significantly. In practice, a is usually specified by the space limitations. Moreover, Q of the Rx coil should be kept greater than 1, not to attenuate the signal. The presence of C_r capacitor is always necessary to adjust the phase of the received signal, such that it can be fed into amplifiers with a resistive input impedance.

3.4 Summary of Coil Design Criteria

- The Tx coil should have the smallest possible a while the Rx coil should have the largest possible a.
- The cross section of the Rx coil should be circular or square (when a circular shape is difficult to manufacture), while the cross section shape of the Tx coil might be freely chosen based on practical winding issues.

  Combining the above two statements suggests that the Tx coil should be single layered with a very large length (c) and $a = b/2$ which is impractical. Therefore, in practice, the maximum permissible length of the Tx antenna (c_{Max}) should specify the coil's dimensions. As for the Rx coil, the maximum permissible radius (a_{Max}) specifies its size.
- The Rx coil center ought to be located on the Tx coil axis. Otherwise the attenuation in $|\mathbf{B}|$ could be approximated by the plot in Fig. 7.5.
- When the coils and their locations are known, the receiver should face the right direction such that $\mathbf{B}$ is perpendicular to the Rx coil surface. In worst case, when $\mathbf{B}$ and the coil's surface are parallel, no signal will be received.

4 Discussion

Having found the best conditions for Tx and Rx antennas, it is time to investigate feasibility of communication in our district heating problem. We have taken the following assumptions into account:

For the Tx antenna, 150 m of wire with $R = 0.3$ mm and electrical resistance 0.06 Ω/m is used to make the coil with $a = 1.30$ cm and $c = 5.00$ cm. This arrangement results

in 1836 turns of wire in 22 layers with $b = 1.32$ cm. For the Rx antenna, the same wire gauge and length is used, but with $a = 18.0$ cm and a square cross section which result in 133 turns of wire with $b = c = 0.69$ cm. The Tx antenna is located at $z = 0$ plane, centered at the origin, while the Rx antenna is located on $r = z$ cone facing the perfect direction. Note that the location $r = z$ is far from the best position. Being closer to the z axis is preferable in practice.

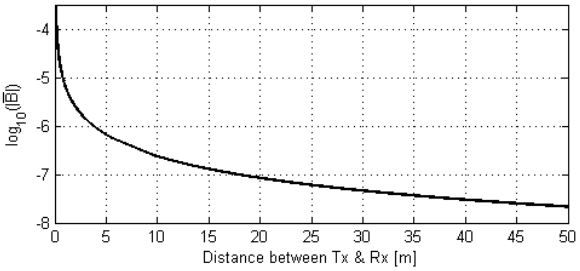

Figure 7.10: $|\bar{\mathbf{B}}|$ of the optimized Tx coil versus distance between Tx and Rx

Electrical current through the Tx coil is assumed to be 100 mA RMS. Fig. 7.10 exhibits $\log_{10}|\bar{\mathbf{B}}|$ as the inter-distance of antennas increases. When the Tx and Rx antennas are 40 m away from each other: $|\bar{\mathbf{B}}| \approx 29.9$ nT.

At the Rx antenna, (7.6) turns out to be 2.50×10^{-2} Hm$^2/\Omega$. It should be multiplied by $\omega^2|\bar{\mathbf{B}}|$ to finally give v_z which is equal to 265.5 mV at the operating frequency 3000 Hz. Assuming 1 mV as the threshold of a detectable signal, the calculated v_z is easy to detect with a standard filter/amplifier combination at the receiver. This ample margin indicates that MI is a working method for making reliable communication links for underground wireless sensor networks.

References

[1] S. A. Meybodi, P. Pardo, and M. Dohler, "Magneto-inductive communication among pumps in a district heating system," in *the* 9^{th} *International Symposium on Antennas, Propagation and EM Theory*, Guangzhou, China, December 2010, pp. 375–378.

[2] S. A. Meybodi, M. Dohler, J. Bendtsen, and J. D. Nielsen, "Feasibility of communication among pumps in a district heating system," submitted to the IEEE Antennas and Propagation Magazine.

[3] R. Bansal, "Near-field magnetic communication," *IEEE Antennas and Propagation Magazine*, vol. 46, no. 2, pp. 114–115, April 2004.

[4] J. Sojdehei, P. Wrathall, and D. Dinn, "Magneto-inductive (MI) communications," in *Oceans, MTS/IEEE Conference and Exhibition*, Honolulu, HI, USA, November 2001, pp. 513–519.

[5] R. Batcher, "Loop antenna for submarines," *Wireless age*, vol. 7, p. 28, 1920.

[6] M. H. Nayfeh and M. K. Brussel, *Electricity and magnetism*. New York: John Wiley & Sons Inc., 1985.

[7] C. A. Balanis, *Advanced engineering electromagnetics*. New York: John Wiley & Sons Inc., 1989.

[8] T. Ishida, H. Ando, and M. Fukuhara, "Estimation of complex refractive index of soil particles and its dependence on soil chemical properties," *Elsevier Remote Sensing of Environment*, vol. 38, no. 3, pp. 173–182, December 1991.

[9] I. Akyildiz, Z. Sun, and M. Vuran, "Signal propagation techniques for wireless underground communication networks," *Physical Communication*, vol. 2, no. 3, pp. 167–183, September 2009.

[10] Z. Sun and I. Akyildiz, "Underground wireless communication using magnetic induction," in *In Proc. IEEE ICC*, Dresden, Germany, June 2009.

[11] ——, "Magnetic induction communications for wireless underground sensor networks," *IEEE Transactions on Antenna and Propagation*, vol. 58, no. 7, pp. 2426–2435, July 2010.

[12] K. Kuns, "Calculation of magnetic field inside plasma chamber," UCLA, Tech. Rep., August 2007.

[13] J. C. Maxwell, *A treatise on electricity and magnetism*, ser. Clarendon press. Oxford, 1873, vol. 2.

Paper C

Feasibility of Communication among Pumps in a District Heating System

Soroush Afkhami Meybodi, Mischa Dohler, Jan Dimon Bendtsen, and Jens Dalsgaard Nielsen

This paper is accepted for publication in:
IEEE Antennas and Propagation Magazine
Expected publication date: August 2012

Abstract

The tackled problem is: selecting a viable method for communication among pumps in a district heating system, viewed as a metropolitan wireless sensor network whose nodes are confined underground and physically connected by pipes. In a further horizon, providing sophisticated control systems for similar urban utilities motivates this research problem. Here, we have reported the results of investigating several potential methods for realizing the idea of "The Talking Pumps" in a district heating system. This includes a diverse list of key references, simulations for some methods, and experimental results for some others, followed by selection the most appropriate option. The considered methods are using 1) Acoustic waves through water and pipelines, 2) Power line communications, 3) Electrical conductivity of pipes, 4) Cell phone infrastructure, 5) Free and guided Radio Frequency (RF) Electromagnetic (EM) waves, and 5) free and guided Very Low Frequency (VLF) electric and magnetic fields, also known as Magnetic Induction. The viability of the latter is verified by simulations and primitive experimental results.

1 Introduction

Advances in the fields of Wireless Sensor Networks (WSN) and Networked Control Systems (NCS) have created many new opportunities for joint designs where a control system routes its required signals via wireless communication links. Specific applications are emerging among urban utilities and infrastructure e.g. water distribution systems [1, 2, 3] and district heating systems (DHS) [4]. Unlike electricity distribution, another urban utility, that needs a sophisticated dispatching control with access to real time network-wide data to prevent instabilities, the above mentioned utilities may work well by making use of local simple controllers, in expense of fairly low energy efficiency [5]. However, ever-increasing energy prices and environmental concerns have leveraged the technological advances toward more energy saving solutions. Our focus in this paper is on the feasibility of a communication framework that is going to host the control system for a district heating plant.

District heating is a method for providing heat demands to closely located buildings by delivering hot water to them and cycling the return cooled water into heating plants via dedicated pair-wise pipelines. They create a closed pressurized hydronic system in which water flows by centrifugal pumps. Pipes are buried underground and pumps are usually placed in the basement of buildings along the pipes. They are wired to local pressure sensors to provide local control of the hydronic dynamics. In the traditional design (Fig. 8.1), there are a few pumps for many end users.

Assuming a minimum pressure across all end users is required, each pump should be operated based on the differential pressure across the farthest end user [6]. Measured pressure of the farthest end user is currently transmitted over internet at prescribed time intervals. A human operator uses this data besides weather conditions and weather forecast, to roughly change the pump operating settings. A drawback of such a control method is the considerable loss of pumping energy due to presence of pressure reducing valves across end users which are located close to the pump. Furthermore, the diameters of pipes are optimized and chosen large enough to prevent pressure loss in far loads. This increases heat loss due to large pipe diameters and deteriorates overall energy efficiency.

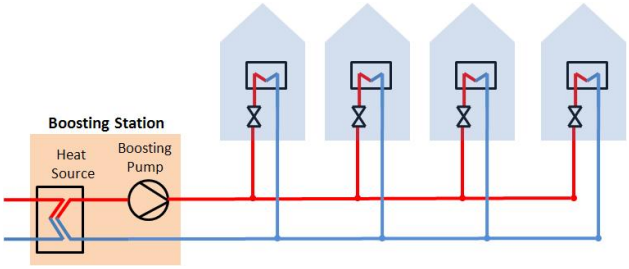

Figure 8.1: District heating piping layout: The traditional design

In a newly proposed piping layout [5] which is designed to improve energy efficiency, the number of pumps are increased and cooperation among them is mandatory to maintain stability of the hydronic network (Fig. 8.2).

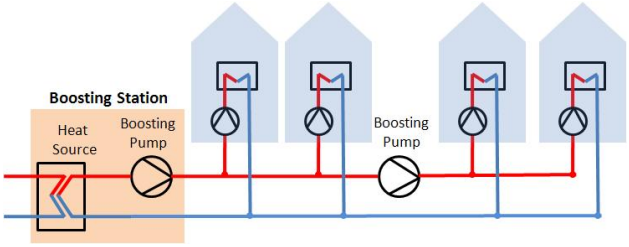

Figure 8.2: District heating piping layout: New design

The goal is to divide the burden of the big old pump among smaller new pumps in an optimal manner. The new pumps are to be placed at each end user and in boosting stations. This layout results in saving pumping energy and facilitates the use of pipes with smaller diameters in future installations, hence contributes to heat loss alleviation. On the other hand, this structure requires data communication among pumps, e.g. data from end user pumps must be sent to their upstream boosting pump [7].

Signaling among the pumps is not trivial. They are located in buildings' basements, which are typically air-filled underground chambers with concrete walls. Pipes are already laid and there is no chance to add wires along them to connect them together.

In a previous paper [8], we investigated the required coverage range of a host wireless network by extracting the empirical cumulative distribution function (CDF) of pumps' inter-distances. We found that a range of 40 m provides at least one in-range neighbor for each pump, while 60% of the pumps have 5 accessible neighbors. Although it does not guarantee full connectivity for the network graph from mathematical point of view, it gives acceptable results in commonplace piping layouts. Therefore, our goal in this paper is to choose and verify a communication method for a typical 40 m through-soil range of an underground network.

Several methods have been tried in the literature to solve similar problems, but there is still no outstanding solution. Section II introduces these methods. Cited references are in some cases backed by simulations, experiments, and feasibility assessment which

contribute to prove inadequacy of the inspected methods.

The main focus of the paper is on *Magnetic Induction (MI)* which is introduced in Section III. MI has already found several applications. It is used in very low power wireless microphones and headsets that are used by security forces [9, 10, 11] in frequency ranges 11-15 MHz and distances up to 3 m providing a high level of security by involving quasi-static fields - to be described later. Lower frequencies of 300 Hz to 4 KHz are employed for distances up to a few kilometers in soil or water medium. They have been used in underground and underwater voice and data communications for mining and military purposes, underwater navigation, remote triggering of underground and underwater mines and ammunition, seismic fence, and submarine communication systems [12, 13, 14, 15]. We will see later in this paper, that in such cases the communication media are hardly penetrable by higher frequencies.

Section IV contains a simulation study on optimal design of MI transceiver antennas. It is followed by experimental results in Section V which confirm viability of MI for making underground communication links of the required range. Section VI concludes the paper.

2 Various Communication Technologies for District Heating

2.1 Radio frequency electromagnetic waves

This method makes use of Industrial, Scientific, and Medical (ISM) unlicensed radio bands and employs them for underground communications. In contrast to above-ground systems, the main challenge in *Physical* layer is coverage rather than interference mainly because of the high propagation losses of EM waves through the soil [16].

2.1.1 through-the-soil communications

The maximum transmission range in RF systems depends on transmission power, antenna gains, propagation losses and receiver sensitivity. Propagation losses in a homogenous environment consist merely of pathloss which depends on dielectric properties of the medium and distance between transmitter and the receiver. Dielectric properties of soil depend highly on the soil composition and water content [17, 18] that varies quite a lot for different locations and from time to time. This results in a large uncertainty in pathloss behavior besides its typical large value.

Reference [19] gives insight to the maximum achievable transmission range with different frequencies, transmission powers and volumetric water contents of soil. It is shown that the range can be extended only up to 4 m with 10^{-3} BER for nodes buried in a depth more than 2 m, assuming operating frequency of 400MHz, 30 dBm power transmission, volumetric water content of 5%, and PSK modulation which are all quite good conditions. More water content and higher frequencies exacerbate the result. However, this range is certainly inadequate for our problem.

2.1.2 Combined underground above-ground

The nodes in our network are not buried, but placed in the building basements. This will provide the chance for creating multiple paths for a signal, through walls from the base-

ment to the surface, transmission above the surface, and through walls to the basement again. In an unrealistic ideal condition depicted in Fig. 8.3, i.e. with the existence of ideal reflecting surfaces above ground, the wireless signal should penetrate at least through four thick walls.

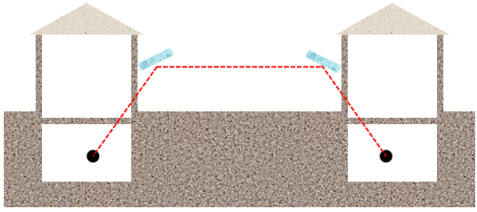

Figure 8.3: A typical underground above-ground combined scenario

Based on the empirically derived pathloss models in [20] which are valid for the operating frequencies of 2-6 GHz, we have estimated pathloss as a function of horizontal distance between the transmitter and the receiver for a set of frequencies. The arrangement in Fig. 8.3 is similar to the indoor room-to-room scenario in [20] with four thick walls.

For this scenario, pathloss (PL) can be estimated as follows:

$$PL[dB] = \quad 20\log_{10}(d[m]) + 46.4 + 20\log_{10}(\tfrac{f[GHz]}{5.0}) + 12n_w \tag{8.1}$$

where d is the distance between transceivers, f is the operating frequency, and n_w is the number of thick walls [20]. The results are exhibited in Fig. 8.4.

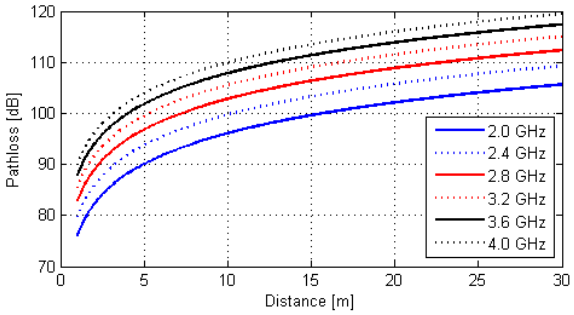

Figure 8.4: Pathloss vs. horizontal distance for several frequencies

Figure 8.4 suggests a maximum transmission range of 15 m in ideal conditions, assuming a maximum tolerable signal loss of 100 dB, which is the dynamic range of current embedded radios. It is still less than the required range of 40 m, but it also shows that there could be a chance if lower frequencies are employed. However, (8.1) is not valid for f ¡2.0 GHz.

To see if lower frequencies can be useful, we ran a test with APC220 radio modules tuned at 434 MHz. The selected frequency is the lowest unlicensed frequency that could be supported by APC220 radio. The sensitivity of this radio is -112 dB at 9600 bps, and it uses GFSK modulation.

The experiment was a simple broadcast test. The transmitter was fixed in a basement, broadcasting a life signal once every second. The receiver, which was a mobile node, was listening to the life signals. We started with both nodes placed close to each other, then got out of the basement with the receiver node and moved towards the neighboring basement located in 10 m distance. It turned out that as soon as we were to go down in the second basement, the life signal was completely lost. This result is due to the lack of the required reflecting surfaces in real scenarios. We did not pursue this approach any further.

2.1.3 Pipelines as waveguides

Another idea is to combine the ubiquitous embedded radios with the potential waveguide properties of the pipes. The question is whether a water filled pipe can act as an efficient waveguide. Although it has already been investigated with air filled metallic ducts with some degrees of success [21], there is a skepticism regarding water as the inner material. An EM waveguide principally consists of either two dielectric materials, i.e. a *dielectric waveguide* or a metal and an insulator, i.e. a *matallic waveguide*, but if none of the materials is an insulator, the wave cannot be propagated through the tube.

Water, as is used in district heating, is neither a good conductor, nor a good insulator. Its conductivity is less than 10^{-3} S/m [6] compared to 2.0×10^{6} S/m for steel [22] and 0.3×10^{-14} to 0.8×10^{-14} S/m [23] for air. These numbers motivated us to do a simple experiment to measure the attenuation in a water filled pipe presumed as a waveguide.

Although it has been subject to special treatments e.g. de-aeration and removing some ions and particles, its conductivity is in order of $< 10^{-3}S/m$ [6] compared to 5×10^{-4} to 5×10^{-2} for fresh drinking water, $2.0 \times 10^{6}S/m$ for steel [22] and 0.3×10^{-14} to $0.8 \times 10^{-14}S/m$ [23] for air. These numbers motivated us to do a simple experiment to measure the attenuation in a water filled pipe presumed as a waveguide.

The experiment setup is shown in Fig. 8.5. We used a 2.5" iron pipe conforming to DIN2440 standard. It was integrated into the potable water distribution pipelines. As the transmitter and the receiver, we used coil antennas each made of 83 turns of single layered insulated copper wire conforming to AWG 30 standard. Lengths of the coils turned out to be 40 mm. Both antennas were shielded in cardboard boxes which were totally covered by a thick aluminum foil. Therefore, the signal could only transmit through the pipe.

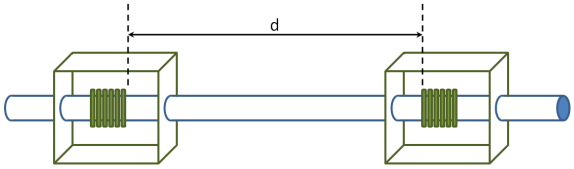

Figure 8.5: Metallic waveguide experiment setup

The coils were used with a *Signal Generator* and *Spectrum Analyzer* without impedance matching circuits. Fig. 8.6 shows signal loss as a function of distance (d) between the coils for operating frequency 400 MHz. Impedance mismatch of the two circuits imposes a large constant signal loss, but we are generally interested in the variations of signal

loss versus d. Therefore, it is possible to ignore the matching problem assuming that the matching parameters are not affected considerably while d varies.

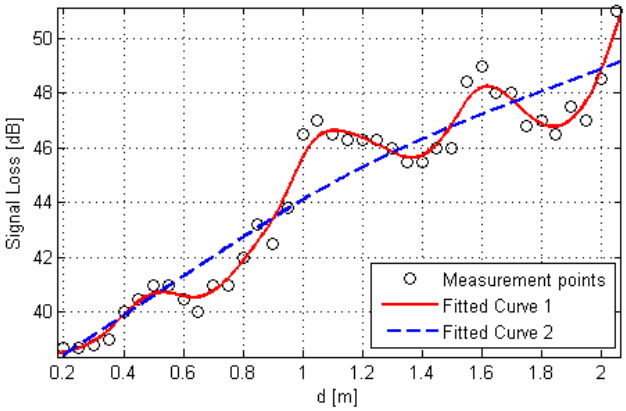

Figure 8.6: Waveguide effect on signal loss at 400 MHz

The oscillatory behavior of signal loss admits existence of the waveguide effect. The observed wavelength is approximately 0.5-0.6 m which is in accordance with the calculated wavelength for 400 MHz in water medium, i.e. 0.56 m, provided that:

$$v = \frac{c}{n} \qquad (8.2)$$
$$\lambda = \frac{v}{f}$$

where c (2.9979×10^8 m/s) is the speed of light in vacuum, n (1.333) is the *index of refraction* of water, v is the speed of light in water, λ is the wavelength, and f is the operating frequency [22].

Measured data are smoothed and plotted in Fig. 8.6 in two steps. The first fit explicitly shows the oscillatory behavior, while the second fit shows the general trend and anticipates, via extrapolation, 100 dB of signal loss at approximately 20 m after impedance matching. Unfortunately, we did not have access to longer straight water filled pipes to verify this result, but it is not satisfying for our DHS problem yet, considering the required range of 40 m and the presence of bends and branches in pipes, in practice, which severely affect the waveguide properties.

2.2 Acoustic waves

Another intuitive option to use the infrastructure itself as the communication medium is to employ the mechanical vibration in form of sonic/ultrasonic waves as the carrier of data. Either pipes or water could be used as the main medium. There is an intrinsic advantage in using water because there could be variations in the material of the pipes in some cases in practice. However, we have considered both cases. Most of the times, pipes are made of steel and have welded fittings [6], but sometimes plastic pipes are used for smaller end user connections. Making use of pipes creates extra dependencies between the WSN

and the DHS, making both of them less flexible for future design modifications simply because the pipes are required to fulfill more criteria. Nevertheless, we will consider both pipes and water. In order to extend the results of this section to water distribution systems, one should bear in mind: 1) the dependency of sound speed in water to the water temperature and salinity, if water is used as the communication medium; and 2) the effect of the different piping fittings i.e. welded versus threaded or flanged, if the pipelines provide the channel.

2.2.1 Acoustics through the water

Using water as the main communication medium has received less attention compared to in-pipe propagation. Sound speed in water is about 27% of in iron, i.e. 1500 m/s versus 5500 m/s [24]. Therefore, the wavelength at the same frequency is shorter with the same proportion, making the wave more prone to destructive multipath effect. *Doppler Effect* is also another issue since the speed of water is comparable with the speed of sound [25].

The most successful application in this area is underwater open-sea communication where the multipath effect is minimized. Usages are including oceanography, environmental monitoring, *tsunami* warning, etc. [26] with commercial products.

Another working example which is more relevant to in-pipe communication is *mud pulse telemetry* [27]. It is used to provide measurement while drilling in oil and gas industries. While a pump at the surface pumps mud into the borehole and the pressurized mud drives the drill-head, the drill-head's built-in valves produce pressure pulses against the mud flow. The energy to create such pulses comes from the pressurized mud itself. These pulses are then received at the surface and the data are extracted. The typical commercial data rate is 12 bps for boreholes shallower than 7000 m. The highest reported data rate is about 40 bps for shallow boreholes down to 3 bps for boreholes deeper than 11000 m [28].

2.2.2 Acoustics through the pipes

Besides pipe's material and geometry, propagation of sound waves in a pipe depend on the materials inside and outside of it. Simplest results are achieved if there is vacuum inside and out. Simulation and experimental results are reported for gas filled pipes in [29]. The application is remote reading of meters for natural gas distribution companies [30]. A two-way data transmission rate of 2560 bps is reported with 1×10^{-2} bit error rate (BER) for transmission via a typical six floor building with branches and bends in pipes, assuming in-line speakers and microphones. However, no commercial products were traced at the time of writing this article.

Wave propagation in water filled pipes are investigated in [24], [31], and [32]. The second chapter of [31] offers a clear and compact literature review on the subject, while [24] gives a detailed analysis of the dispersion of different acoustic modes in a water filled pipe. Open-sea and in-pipe cases are compared in [32]. All of these references state that the main source of signal attenuation in pipes is the reverberations which depend heavily on the geometry, fittings, bends and branches of the pipes that could be completely different from one case to another. It considerably reduces the achievable data rate by adding extra echoes, pushing us to use lower frequencies and requiring more sophisticated signal processing.

The most successful relevant application is pipe leak detector [32] which is not a two way communication system and requires a low data rate. In such systems, microphones are installed on pipes with prescribed distances. They listen for a continuous noise which will be interpreted as the aftermath of water leakage from a pipe breakage. Provided that there is a synchronizing facility among microphones, the location of the breakage can be calculated by measuring the differences of time-of-arrivals of signals. There are commercialized cases on the market.

Experimental results [32] for steel pipes with the diameter of 10 cm, predict the uncoded BER 32% for data sent at the rate of 7 kbps at 21 kHz with FSK modulation via water in a straight buried pipe. The transmitter and receiver are located approximately 9 m away from each other. Although the BER could be remedied by signal processing techniques, the transmission range is obviously insufficient. Furthermore, the microphone and speaker which were used in this experimental study were installed inside the pipe and precisely centered at the pipe axis. However, in practice, clamped equipments are preferred than flanged or threaded ones due to feasibility and maintenance considerations.

From financial point of view, a general survey on the commercial microphones, hydrophones and speakers show that there are some contact microphones and speakers that can be clamped on pipes, but they are not originally designed for in-pipe use. Therefore, the performance is not tested. Moreover, the price is comparable with a DHS small pump.

To get an understanding of the performance of the economical options, we performed a simple test with fairly cheap piezoelectric buzzers used as contact microphone and speaker. The buzzers were fixed with epoxy glue on an iron air-filled open-ended pipe. Single carrier transmission test at 4.1 kHz, which was the piezoelectric resonance frequency, showed that a maximum 10 v input voltage at the transmitting buzzer produces less than 0.1 mv at the receiving buzzer at a distance of 0.5 m. Although the sonic signal could be heard clearly, it was not detectable by the oscilloscope for larger distances. In summary, it was understood that a microphone with a high sensitivity, should be used and it costs too much even in bulk numbers.

Last but not least, the communication framework that we are interested to build is going to host a control system for regulating pressure and compensating pressure disturbances in a DHS. This means that the transmitted data have the highest importance when pressure irregularities happen, but these irregularities also induce vibrational noise which degrades reliability of the communication link. Therefore, the channel is least reliable when it is most needed.

Marginal reported experimental results for short pipes, the high price of the required hardware, i.e. comparable to the price of the pump itself, and the effect of the process disturbance on the communication channel makes the current acoustic technologies unreliable for in-pipe communications.

2.3 Electrical conductivity

2.3.1 Power line communications

Power Line Communications (PLC) makes use of the power grid as a wireline network for data transfer. The idea has been around since 1920's for voice communication, telemetry, and supervisory control [33]. The technology was revisited two times: 1)in late 1970's, motivated by the advent of digital signal processing, in order to be used as a means of load

management and remote meter reading in electrical networks [34]; and 2)in late 1990's to host broadband services like internet over power lines [35].

The main drawback of using PLC is the variety of applications and non-interoperable standards which are thriving and competing to gain access to the readily available power grid infrastructure. Such a competition has become harsher than ever by the advent of Smart Grid. A comprehensive and recent study on applications and standards can be found in [36]. Urban utilities, including electrical power distribution, water distribution, gas distribution, and district heating are not considered so far to use PLC in any application other than automatic metering. Therefore, finding space for bandwidth demanding applications like real time control of such plants involves lots of lobbying and non-technical settlements.

There are some technical difficulties in relying on PLC too. Frequency ranges in PLC start from a few hundred hertz for long distance narrow-band applications up to 200 MHz for home applications with a very short range and high bandwidth. The power grid is designed for 50-60 Hz which raises EMC considerations and mutual interference for employment of higher frequencies, especially for those more than 80 MHz due to the presence of ubiquitous TV broadcast channels. This fact has limited the permissible transmission power and has put off the standardization processes.

Another problem is the existence of transformers in power networks which filter out high frequencies. Therefore, bypass devices are needed to be installed parallel with transformers [37]. It explains why PLC has been mostly successful in home automation and intra-building control systems where there is no electrical isolation; and low transmission powers can still cover the required range.

In a DHS, nodes are geographically dispersed and could be located in areas supplied by different transformers. Without the use of bypass devices, a partitioning is imposed on the network which is probably not in agreement with the control system structure. If we accept this major limitation, PLC is a feasible low bandwidth solution, at least from the technical point of view, but maybe not a timely solution due to lack standards for our specific application and interference with other existing systems.

2.3.2 Pipes as wires

This is another intuitive and cheap solution, provided that the pipes are all metallic and electrical conductivity is preserved across the fittings. The idea is to use the conductive pipes as electrical wires. There is only one major problem, i.e. Grounding.

In water distribution systems, pipes are in contact with the soil either directly or through their thin layered coating. This forms a frequency dependent electrical resistance (R_{leak}) between the pipe and earth, as the electrical ground. Large R_{leak} is desired at the operating frequency. There is also another frequency dependent electrical resistance (R_{pipe}) between two neighboring nodes on the pipes. It is shown that with an appropriate use of the skin effect, by choosing the right operating frequency such that R_{leak}/R_{pipe} is maximized, it is possible to make use of this leaky multi-ground system to send and receive signals [38].

However, in a DHS, pipes are heavily insulated with foam-filled PVC jackets in order to prevent heat loss. Therefore, they are required to be grounded with copper wires as they enter each basement to hamper electrostatic shock hazards. Moreover, the following formula [22] shows that the skin depth of steel (pipes' material) is about 10 to 20% of

copper (earth wire).

$$\delta = \sqrt{\frac{\rho}{\pi f \mu}} \tag{8.3}$$

In (8.3), δ is the skin depth, ρ the conductivity, f the frequency, and μ is the magnetic permeability of the conductor which varies significantly for different kinds of steel. Therefore, increasing the frequency will even aggravate the attenuation by increasing $\frac{R_{pipe}}{R_{leak}}$.

In summary, dependency on pipes' material, direct grounding of pipes, and the stronger skin effect for pipes rather than earth wires, diminish any chance to transfer signals through the pipe by this method.

2.4 Cell phone infrastructure

There are lots of commercial products and some publications which claim to offer remote monitoring and control of home appliances by mobile phones. The idea is to use two SIM cards and transfer data by Short Message Service (SMS). It works well, when a single measurement or command needs to be sent. Switching home appliances ON and OFF, sending an alarm signal from the field to a remote operator, or automatic reading of electricity/water/heating meters once a month/year are among success stories, but it cannot be applied to our closed loop control problem due to the following concenrs.

First of all, the system will be dependent on a third party infrastructure which is not desirable. Second, in every closed control loop, signaling should be fast enough compared to the dynamics of the measured variable, i.e. water pressure in our case. In hydronic systems, changes in water pressure propagate with the speed of sound in pipes, which is 1500 m/s. Comparing this speed with typical distances in an urban area suggests a sampling rate faster than 1 Hz. Sending a SMS at every second is quite expensive, if possible. Third, from technical point of view, SMS is sent via signaling paths which are originally provided to control the telephony traffic. They are separate from main voice path and have a maximum payload capacity of 1120 bits. The signaling paths let SMS go through whenever no higher priority traffic control signal exists. Therefore, congestion can happen resulting in a varying delay of several seconds. SMS centers store and retry to forward old messages by default. This function exacerbates the situation by making a long queue at the SMS center. Moreover, reordering of the gathered data may happen unless at least a local time stamp fills a portion of the payload and is stored in the receiver.

Furthermore, our application requires the nodes to be placed in the basement of buildings where signal strength is greatly attenuated even for well covered cellular networks. Based on authors' observations, even in automatic reading of meters which is a quite successful application for cellular monitoring, regular human operator check is required when the apparatus is located in the basement.

3 Magnetic Induction

MI has been around since 1920's being used as a means of communication between submarines [14]. There is no fundamental difference between MI and RF EM waves. However, because of different technical specification, most importantly antenna size, available bandwidth and transmission power, various names have popped up to describe it including

Very Low Frequency (VLF), *Ultra Low Frequency (ULF)*, and *Extremely Low Frequency (ELF)*. Each of these names addresses a certain range of frequencies which are not even understood in the same way, in different communities. For example, *National Aeronautics and Space Administration (NASA)*, *World Health Organization (WHO)*, and *International Telecommunication Union (ITU)* have all assigned different frequency ranges to define the above terms. In this article, we use MI as the only alternative term and explain why it is worth being distinguished from usual radio spectrum.

MI also employs alternating Electric ($\mathbf{E}$) and Magnetic ($\mathbf{B}$) fields as data carrier, but it uses low frequencies and low power such that the communication range falls off within the distance predominated by the *static* or *quasi-static* fields, as described in the following.

$\mathbf{E}$ and $\mathbf{B}$ fields which are created by an alternating electrical current in an antenna can be estimated by the summation of several terms that have different dependencies to distance *r* from the antenna, e.g. $|\mathbf{B}| \approx \alpha.1/r^3 + \beta.1/r^2 + \gamma.1/r$ in which α, β, and γ depend on the antenna pattern and its driving current. In the vicinity of the antenna, fields are very complex (to be shown later in Section IV). Gradually, they are predominated by $1/r^3$ term. Up to this zone, is called the *static field* domain. Other terms predominate at farther distances and are called *inductance field* (*quasi-static field*) and *radiation field*, respectively [22]. The latter accounts for EM radiation in which $|\mathbf{B}| \propto 1/r$ and $|\mathbf{E}| \propto 1/r$. To define which field component is predominant, r should be compared with *wave number, k*:

$$k = \omega/v \tag{8.4}$$

where ω is the angular frequency of the $\mathbf{E}$ and $\mathbf{B}$ fields and v is speed of EM waves in the medium. $k.r \gg 1$ and $k.r \ll 1$ imply that either the radiation or the static field is dominant, respectively. Otherwise, the field is quasi-static.

Static and quasi-static fields decay much faster than the radiation field while getting away from the current source. More concretely, their energy flux *Poynting vector* decays faster than $1/r^2$ and the emitted power per area tends to zero as distance from the source goes to ∞. Therefore, they are non-radiating fields and have no value in long range communications. However, if the frequency is low enough, the distances of interest can still be perceived as short distances. Hence, even non-radiating fields can be exploited to transfer energy between transceivers. In our underground communication problem, the choice of the operating frequency leaves us in non-radiating field domain.

Why should we choose such low operating frequencies? The answer lies in the dielectric properties of lossy materials like wet soil. When materials are influenced by alternating $\mathbf{E}$ and $\mathbf{B}$ fields, their electric susceptibility (χ_e) can be stated as a complex function of frequency:

$$\chi_e(\omega) = \chi'_e(\omega) - j\chi''_e(\omega) \tag{8.5}$$

The general behavior of real part $\chi'_e(\omega)$ and imaginary part $\chi''_e(\omega)$ for most of the materials is sketched in Fig. 7 [39]. $\chi'_e(\omega)$ increases most of the times with respect to ω, except for short intervals where it drastically decreases. These intervals coincide with local maxima in $\chi''_e(\omega)$. $\chi'_e(\omega)$ defines velocity of $\mathbf{E}$ and $\mathbf{B}$ fields in a matter, while $\chi''_e(\omega)$ accounts for attenuation due to molecular resonance phenomena.

It provides an extra *alternating field conductivity* (σ_a) besides the common *static field conductivity* (σ_s) at resonance frequencies. The overall conductivity (σ_e) can be stated as

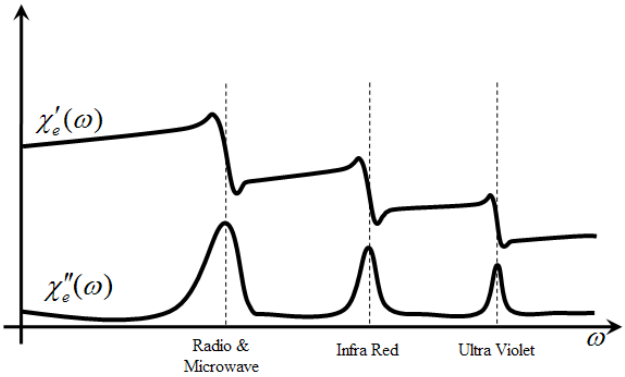

Figure 8.7: Electrical susceptibility of matters versus frequency

[39]:

$$\sigma_e = \sigma_s + \sigma_a = \sigma_s + \omega(\epsilon_0 \chi_e''(\omega)) \tag{8.6}$$

Equation (8.6) and Fig. 8.7 imply that, $\mathbf{E}$ and $\mathbf{B}$ fields should have frequencies lower than the common radio and microwave bands, i.e. the first peak in $\chi_e''(\omega)$ graph, in order to decrease overall conductivity, hence attenuation. Furthermore, in non-radiating fields, there holds no relation such as $|\mathbf{E}|/|\mathbf{B}| = v$. Therefore, $\mathbf{B}$ can be substantially large. It also means that the energy does not have to be divided equally between the $\mathbf{E}$ and $\mathbf{B}$ fields, as it does in radiating fields.

A starting point to choose the appropriate operating frequency is to compare our problem with other MI applications. Knowing that the *index of refraction* (n) of EM waves in soil is $\approx$ 1.5 [40], the speed of EM waves in soil is given as $v = c/n \approx 2{\times}10^8$ m/s. Therefore, for distances of 10 m to 70 m and working frequencies of 2 kHz to 9 kHz, we have $k.r \in$[0.006,0.01] which falls in the non-radiating field domain and does not interfere with any licensed frequencies.

3.1 System Components and Circuits

Fig. 8.8 shows the basic elements of a MI communication system. Coil antennas are used at the transmitter (Tx) and receiver (Rx) sides. Coils have low radiation resistance and transmit very little real power. Therefore, they have low performance for radiation field. Instead, a coil antenna has a high reactive power going back and forth to the antenna and the surrounding space in each cycle. This is a desired property for MI, unlike typical RF systems in which reactive power is kept as low as possible.

In comparing transmission power of MI and radio systems, one should note that unlike radio systems in which the consumed real power is transmitted and usually measured in dBm, a MI transmitter does not transmit real power. The consumed real power in a MI antenna is the wasted power in R_t as it is converted to heat. For instance, a commercial MI system might have a nominal power consumption between 100 W and 200 W which will be confusing if carelessly transformed into dBm scale.

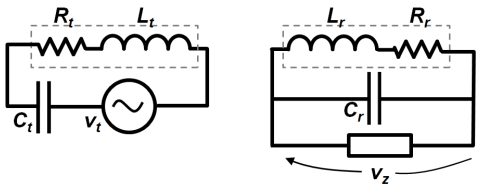

Figure 8.8: MI transmitter and receiver circuits

Two coil transceivers operating at low frequencies can be viewed as a core-less transformer with two separate coils and inefficient coupling. Both coils can be fairly modeled as an ideal inductor in series with a resistor which is equal to resistance of the coil wire. We assume that the operating frequency is low enough, hence ignore the self capacitance across the coil in its lumped model. However, resistance of the coil wire depends on the operating frequency through *skin* and *proximity effects*.

At Tx side, the objective is to produce the strongest possible $\mathbf{B}$ field. This will be achieved when electrical current in the coil (I_t) is maximized. Therefore, a capacitor C_t should be connected to the coil in series and bring it to resonance such that $I_t = v_t/R_t$. In practice, when R_t is small, a voltage source with low output impedance should be used to be able to provide the required current.

In order to increase I_t further, R_t should be reduced. It can be achieved by breaking the coil into several parallel coils, connecting some/all of the individual wire turns in parallel [41]. If the total DC resistance is R and the wire turns are to be grouped in N parallel coils, each with DC resistance of R/N, the overall resistance (R_t) will be equal to R/N^2.

Increasing I_t will increase the voltage across C_t which is equal to v_t transformed by the *quality factor* of the Tx coil. In practice, when I_t is in order of tens of milli-amperes, the voltage across C_t could rise up to tens of volts. Therefore, C_t should have a high breakage voltage. As we will see later in this paper, the maximum permissible voltage for C_t was the parameter which limited I_t in our experiments.

The Rx coil, on the other hand, should be sensitive to the slightest change of the passing magnetic flux and produce the highest back-emf. The induced RMS emf v_r can be calculated in a multi-turn coil by:

$$v_r = \omega . \sum_{i=1}^{N_r} \left[\int_{A_i} \bar{\mathbf{B}} . \hat{\mathbf{n}} d(s) \right] \tag{8.7}$$

where ω is the angular frequency of I_t, A_i the area of individual turns of the Rx coil, N_r the number of turns of the Rx coil, and $\bar{\mathbf{B}}$ is the RMS magnetic field passing through the it. If $\bar{\mathbf{B}}$ has a constant amplitude at all the points on the Rx coil surface and is parallel to $\hat{\mathbf{n}}$, which has unit length and is perpendicular to the coil surface, (8.7) will be simplified as follows:

$$v_r = \omega . |\bar{\mathbf{B}}| . \sum_{i=1}^{N_r} A_i \tag{8.8}$$

which shows that $\sum_{i=1}^{N_r} A_i$ should be maximized in order to get the highest possible

back-emf. Furthermore, if a capacitor C_r is connected to both ends of the coil such that it brings the coil into resonance (see Fig. 8.8), the followings hold:

$$\begin{aligned} C_r &= \frac{1}{\omega^2 L_r} \\ v_z &\approx \frac{1}{jC_r\omega} I_r = -j\omega L_r (\frac{v_r}{R_r}) \\ &= -jQv_r \end{aligned} \tag{8.9}$$

provided that the input impedance of the receiver amplifier/filter is very large and the *quality factor* (Q) of the coil is defined as $Q = \omega L_r / R_r$. Therefore, if the Rx coil has a high Q, the passive band-pass receiver circuit acts as a pre-amplifier.

Combining (8.8) and (8.9), the design objective in Rx antenna is to maximize the following expression:

$$\frac{L_r}{R_r} \cdot \sum_{i=1}^{N_r} A_i \tag{8.10}$$

As another approach, coupling between transmitter and receiver antennas can be formulated as mutual inductance between the two coils [19]. In district heating and water distribution systems, it is possible to amplify mutual inductance by employing pipes as ferromagnetic cores of the transceiver antennas. This will increase the self and mutual inductance in expense of adding core losses. We have shown in [8] that mutual inductance can be enhanced by several orders of magnitude when coils are wound around steel pipes. However, iron losses counteract mutual inductance enhancement, thus we have used non-metallic core coils throughout this article.

Employing intermediate coils as magneto-inductive waveguides is another technique to increase the mutual inductance in specific directions and are investigated in [42], and [43]. However, it is not practical in our problem to insert intermediate coils because of the lack of pipes accessibility at regular distance intervals.

4 Coil Antenna Design

Given a specific wire gauge R, and length l, an interesting problem is how to wind the coils such that the design criteria in Tx and Rx antennas are met. Assumptions on wire gauge and length are reasonable since they define cost and weight of the coils. They hold throughout this section.

We start by introducing the analytical expression for the $\mathbf{B}$ field in all points of the space, created by a single turn current loop as shown in Fig. 8.9. The formula is derived by calculating the magnetic vector potential first, and taking the curl of it to find $\mathbf{B}$ [44].

$$\begin{aligned} \mathbf{B}(r,z) = \frac{\mu I k}{4\pi\sqrt{ar^3}} \Bigg[-(z-h)\left(K(k) - \frac{2-k^2}{2(1-k^2)}E(k)\right)\hat{\mathbf{r}} \\ + r\left(K(k) + \frac{k^2(r+a)-2r}{2r(1-k^2)}E(k)\right)\hat{\mathbf{z}}\Bigg] \end{aligned} \tag{8.11}$$

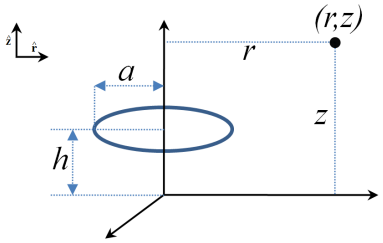

Figure 8.9: A current loop

$\mathbf{B}$ is expressed in terms of complete elliptic integral functions of first and second kind, i.e. $K(k)$ and $E(k)$, where their argument k is defined as:

$$k = \sqrt{\frac{4ar}{(r+a)^2 + (z-h)^2}}$$

All other parameters of (8.11) are expressed in Fig. 8.9. If $\mathbf{B}$ is only sought on z-axis, (8.11) will be significantly simplified. In that case, it is more convenient to directly use Biot-Savart law to calculate on-axis $\mathbf{B}$ field.

4.1 Relative positioning of Tx and Rx antennas

The graphical representation of $|\mathbf{B}|$ for a multi-turn coil is given in Fig. 8.10 as a contour plot at two zoom levels. The coil is not shown in this plot, but it is laid in $z = 0$ plane and centered at the origin, with $a = 0.1$ m.

The field is calculated at each point in space by summing up (8.11) for individual wire turns, considering the exact radius and location for each turn of wire.

It is shown that $|\mathbf{B}|$ is larger around the coil and on the coil's axis. An interesting behavior is observed on $z = 0$ plane where $|\mathbf{B}|$ has a local minimum at a radius of approximately 2.4 a. This is due to the $\mathbf{B}$ field generated by the currents flowing in farther parts of the loop which counteract the $\mathbf{B}$ field generated by the currents flowing in the closer parts. At larger distances from the coil $|\mathbf{B}|$ starts to degrade monotonically.

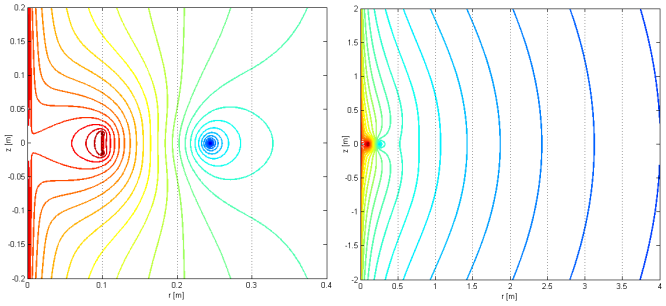

Figure 8.10: Contour plots of $|\mathbf{B}|$ in the space surrounding the Tx antenna at two zoom levels

Since the contours in Fig. 8.10 (right) do not follow an exact spherical pattern, there are variations in $|\mathbf{B}|$ at fixed distances from the Tx antenna. It is clear that $|\mathbf{B}|$ is larger around the z-axis. This behavior is shown quantitatively in Fig. 8.11, in which $|\mathbf{B}|$ is drawn at three different fixed distances of 1 m, 10 m, and 100 m from the Tx antenna. Each curve starts from vicinity of the z-axis, associated to $\theta = 0.1^\circ$ to $z = 0$ plane, associated to $\theta = 90^\circ$. The results are shown in log-log scale in sake of highlighting the differences. Moreover, the curves are shifted vertically such that they all coincide at $\theta = 90^\circ$. Fig. 8.11 shows that $|\mathbf{B}|$ can vary up to approximately 20 dB at the Rx antenna based on positioning of the Tx antenna such that $\theta \in [1^\circ, 90^\circ]$. It also shows that $|\mathbf{B}|$ is more sensitive when θ is small.

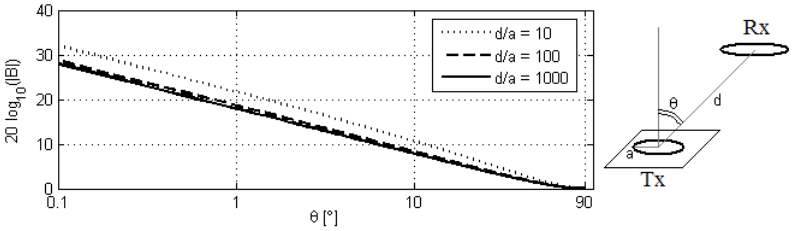

Figure 8.11: Normalized $|\mathbf{B}|$ vs. Rx coil position with respect to Tx coil axis

Note that Fig. 8.11 does not say anything about the angles between Tx and Rx coils' axes, which is important in calculating the magnetic flux passing through the Rx coil.

4.2 Cross section shapes

Given the mean radius of the coil a, and hence the number of its turns N, ($N = l/(2\pi a)$), we have compared rectangular cross section shapes with different side proportions to see its effect on the Tx and Rx antennas. These are ranging from a single layered coil in which $b = 2R$ and $c = Rl/(\pi a)$ to a coil with the maximum number of winding layers which is made when $b \leq 2a$ and $c \geq R^2 l/(\pi a^2)$. The latter holds, provided that each wire occupies a square surface of area $4R^2$ at the cross section of the coil. The most compact wiring, results in hexagonal surface assignment to each wire turn and $c \geq \frac{\sqrt{3}R^2 l}{2\pi a^2}$. All of the above inequalities convert to equalities if the wire is very long such that no hole remains in the center of the multi-turn coil.

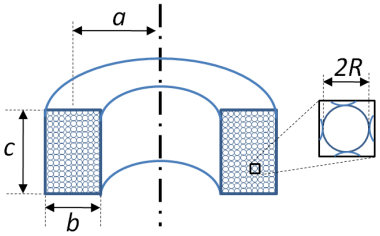

Figure 8.12: Dimensions of a multilayer coil with rectangular cross section

At Tx coil, the difference between $|\mathbf{B}|$ of a single layer and a square coil at $z = 0$ plane is in order of 0.01% of the $|\mathbf{B}|$ created by either of them. In all other points in the space, this trifling difference is even more negligible. Therefore, the cross section shape of the Tx antenna is not tangibly influential on $|\mathbf{B}|$.

At the Rx coil, on the other hand, the coil's cross section shape is quite effective mostly due to its effect on the self inductance of the coil (L_r). This is explicitly presented in (8.10). Since the length and the mean radius of the coil are predefined, $\sum_{i=1}^{N_r} A_i$ is constant and L_r must be maximized.

Maximization of self inductance of a coil was first posed by Maxwell in [45]. He has shown that the *geometric mean distance* between the conductors of an inductor should be minimized in order to get the maximum inductance. This is achieved when the cross section is circular. For rectangular shapes which are easier to manufacture, the best shape is a square, i.e. $b = c$ (See Fig. 8.12). This result is verified by simulations in Fig. 8.13.

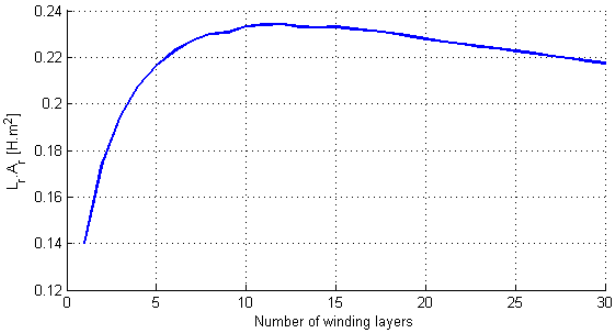

Figure 8.13: emf variations in the Rx coil as a function of cross section shape

The coil in this simulation is made up with 150 m of wire with gauge $R = 0.3$ mm. The cross section is rectangular and the mean radius is $a = 0.20$ m. Consequently, the number of turns N_r turns out to be 119 which yields 11 winding layers for a square cross section. Fig. 8.13 confirms this fact by showing a peak at N_r =11. It also shows that a single layer Rx coil has only 60% performance of the square shaped Rx coil.

In the above simulation, L_r is calculated by injecting a current I into the receiver coil and computing the induced magnetic flux Φ_i in each turn due to the current in all of the turns. Then, $L_i.I = \Phi_i$ gives L_i and $L_r = \sum_{i=1}^{N_r} L_i$. Note that, in calculating the total self inductance of the Rx coil, we only added up L_i of individual turns which convey the self inductance of that turn plus its mutual inductance to all other turns.

4.3 Mean radius vs. Number of turns

Since the total wire length ($l = 2\pi aN$) is predefined, aN is constant and we should make a trade-off between a and N. At the Tx antenna, Fig. 8.14 exhibits the variations in $|\mathbf{B}|$ at $z = 0$ as r increases from 1 m up to 50 m, for different values of a from 0.03 m to 0.25 m.The lowest value is associated with the physical limitation of $a > b/2$ (see Fig. 8.12 for clarification).

Fig. 8.14 suggests that a should be as small as possible to have the maximum $|\mathbf{B}|$ at large distances. The same results were achieved when $z \neq 0$ and also on any straight line

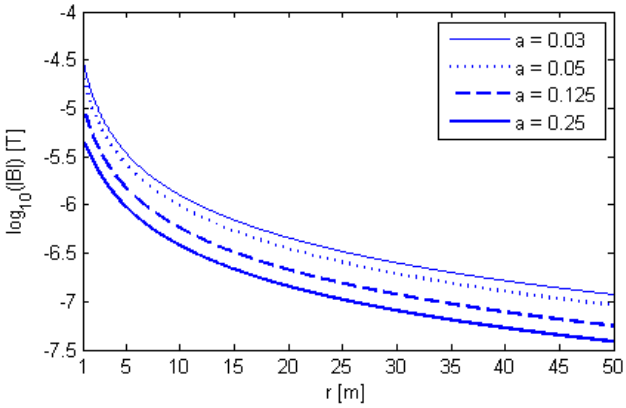

Figure 8.14: Variations in $|\mathbf{B}|$ as r increases

with slope α such that $z = \alpha.r$. No matter how large a finite α is, the above hypothesis holds true and the smallest mean radius gives the maximum $|\mathbf{B}|$ at distances of interest. The only exception is $\alpha = \infty$, i.e. $|\mathbf{B}|$ on the z axis in which the converse is true, but it does not have any practical significance.

For the Rx coil, expression (8.10) is depicted in Fig. 8.15 as a function of a. For all values of a, a square cross section is considered.

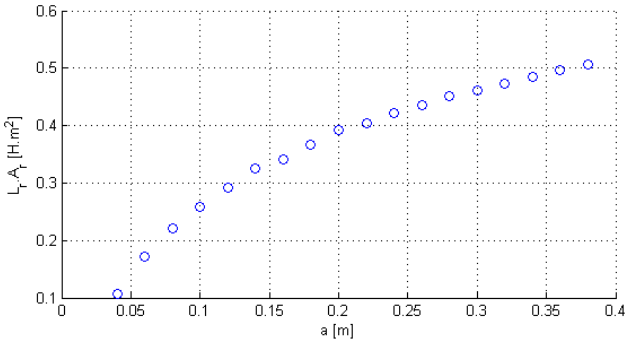

Figure 8.15: Variation in v_z at the receiver coil as a increases

Fig. 8.15 admits that for the Rx coil, larger a produces higher voltages, no matter if the number of turns or L_r is decreased significantly. In practice, a is usually specified by the space limitations. Moreover, Q of the Rx coil should be kept greater than 1, not to attenuate the signal. The presence of C_r capacitor is always necessary to adjust the phase of the received signal, such that it can be fed into amplifiers with a resistive input impedance.

4.4 Summary of Coil Design Criteria

- The Tx coil should have the smallest possible a while the Rx coil should have the largest possible a.

- The cross section of the Rx coil should be circular or square (when a circular shape is difficult to manufacture), while the cross section shape of the Tx coil might be freely chosen based on practical winding issues.

  Combining the above two statements suggests that the Tx coil should be single layered with a very large length (c) and $a = b/2$ which is impractical. Therefore, in practice, the maximum permissible length of the Tx antenna (c_{Max}) limits the size of the coil. As for the Rx coil, the maximum permissible radius (a_{Max}) specifies its size.

- The Rx coil center ought to be located on the Tx coil axis. Otherwise the attenuation in $|\mathbf{B}|$ could be approximated by the plot in Fig. 8.11.

- When the coils and their locations are known, the receiver should face the right direction such that $\mathbf{B}$ is perpendicular to the Rx coil surface. In worst case, when $\mathbf{B}$ and the coil's surface are parallel, no signal will be received.

5 Experimental Results

Having found the best conditions for Tx and Rx antennas, we investigated feasibility of communications in our district heating system. Feasibility study is done by planning a specific simulation scenario and comparing its results with actual measurements from an experiment with the same specifications.

5.1 Specifications and Assumptions

Three coils are used in our simulations and experiments with the following specifications listed in Tab. 8.1. Function of each coil is assigned with respect to the findings in the previous section.

Remark 1: It is assumed that $\mathbf{B}$, unlike $\mathbf{E}$ has a similar attenuation in air and in wet soil due to similar magnetic permeability in these environments. This is explained by the practical separation of $\mathbf{B}$ and $\mathbf{E}$ fields at non-radiating fields domain. Therefore, we have carried out all of the experiments in office environment, not in wet soil.

Remark 2: Operating frequency is chosen to be 5055 ± 1 Hz for all three coils. Variable capacitors are connected to each of the coils and adjusted accurately to attain 5055 Hz resonance frequency. However, since the experiments are carried out in office environment, it is observed that the self inductance of the coils, and hence their resonance frequency is dependent on surrounding ferromagnetic objects, e.g. office heating radiators. Therefore, re-adjustment of the capacitors is needed during experiments.

Remark 3: During experiments, an audio amplifier is used in connection with a digital signal generator at the Tx side to be able to provide the required electrical current. At the Rx side, an oscilloscope receives signals without any additional filter or amplifier.

Table 8.1: Coils specifications

Coil#	Coil 1	Coil 2	Coil 3
Wire length	148 m	359 m	142 m
Wire gauge	0.3 mm	0.35 mm	0.3 mm
Wire diameter incl. insulator	0.65 mm	0.80 mm	0.65 mm
Wire specific DC resistance	0.06 Ω/m	0.045 Ω/m	0.06 Ω/m
Wire measured DC resistance	8.90 Ω	16.19 Ω	8.90 Ω
Coil mean radius (a)	2.8 cm	30 cm	11.3 cm
Coil cross section shape	rect.	circ.	circ.
Coil width (b)	≈3.3 mm	≈12.5 mm	≈10.5 mm
Coil length (c)	111 mm	≈12.5 mm	≈10.5 mm
No. of wire turns	841	190	200
No. of winding layers	5	N/A	N/A
No. of wire turns per layer	169	N/A	N/A
Function	Tx	Rx	Tx/Rx

5.2 Experiment 1 – Measured vs. calculated self inductance

The purpose of the first test was to verify if (8.11) in combination with the procedure introduced at the end of section III.*B.2* can be used to accurately calculate self inductance of a coil. Since only inductance of Rx coils are of interest in our simulations in the previous section, coil 2 and coil 3 are considered here. Calculated values are shown in Tab. 2 against actual values which are measured by two different methods. The first method involved direct use of a RLC-meter to measure L_r. The second method employed $L = 1/(4\pi^2 f^2 C)$ to find L based on using a known capacitor (C) in a LC circuit with our coil and measuring its resonance frequency by oscilloscope. The capacitor was connected in series with the inductor in this test for more accurate measurements in practice.

Self inductance is of interest for the Rx antenna. Thus, the second and the third coils with larger mean radii are considered in this experiment. We have calculated their self inductance by the procedure introduced at the end of section IV-*B*. The results are shown in Tab. 8.2 against the actual values which are measured with two different methods. The first method involves direct use of a RLC-meter to measure L_r, while the second one enjoys a higher accuracy by measuring the capacitor at a specific resonance frequency and employing $L = 1/(4\pi^2 f^2 C)$. The capacitor should be connected in series with the inductor in this test.

Table 8.2: Measured versus calculated self inductance

Coil#	Calc. L	Direct Meas. L	Indirect Meas. L
2	55.11 mH	54.5 mH	54.44 mH
3	18.37 mH	18.17 mH	18.12 mH

The results show only 1% difference between calculated and measured values which is satisfying.

5.3 Experiment 2 – Effect of Surrounding Ferromagnetic Objects

In this test, it was initially intended to verify that the received voltage across the Rx antenna (v_z) is a linear function of the electrical current in the Tx antenna (I_t) in practice, as it is evident from the Biot-Savart law. However, our observations suggest a different title for this experiment.

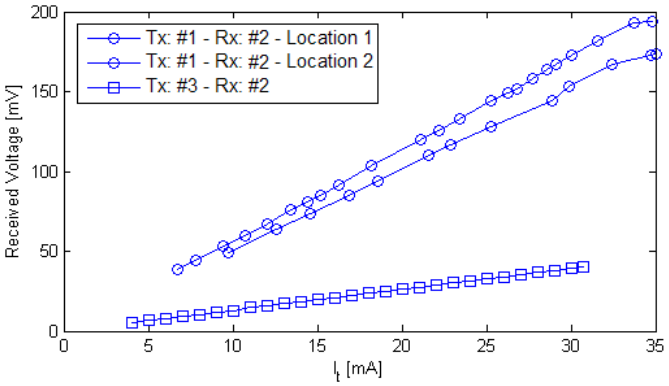

Figure 8.16: Received voltage as a function of Tx current

The test was planned for two combinations where either coil 1 or 3 are used as the Tx and coil 2 as the Rx antennas. Fig. 8.16 shows linear dependency between v_z and I_t for both cases. But the results were not easily reproducible. They varied if nearby ferromagnetic objects were replaced, e.g. office chairs with iron parts. To visualize this effect, two graphs for coil 1 as the Tx antenna are plotted. Placement of the surrounding objects is quite different in one case from the other. All other parameters are unchanged. It can be seen that the slope of the line is changed. This experiment raises some concerns about the effect of steel pipes, in district heating systems, on the $|\mathbf{B}|$ field and self inductance of coils. More experimental studies or real scale implementations are needed to see whether this effect acts in favor or against our system in overall. The bottom line is that various physical piping layouts will impose an uncertainty in quality of the received signal.

5.4 Experiment 3 – Measured vs. calculated back-emf

This is the main experiment, in which we compared simulated and measured values of v_z. Based on (8.9) and backed by the promising results of Experiment 1, this test actually compares simulated and measured values of $|\mathbf{B}|$ at Rx side.

For coil 3 as the Rx antenna, (8.10) turned out to be 1.78572×10^{-2} H.m^2/Ω. This value multiplied by ω^2, where the operating frequency is 5055 Hz, gives 1.80142×10^{7} V/T which is voltage gain of the Rx coil. It is equal to the RMS of v_z if assuming: 1) a uniform $|\mathbf{B}|$ on surface of the Rx antenna, and 2) the ideal direction of the Rx antenna such that its surface is perpendicular to the $\mathbf{B}$ field.

Simulations and experiments are performed on $z = r$ cone. I_t is chosen such that peak-to-peak voltage across C_t does not exceed 40 V, i.e. the maximum permissible voltage for our capacitors. I_t turned out to be 31.0 mA and 28.8 mA RMS for coil 1 and coil 3, respectively.

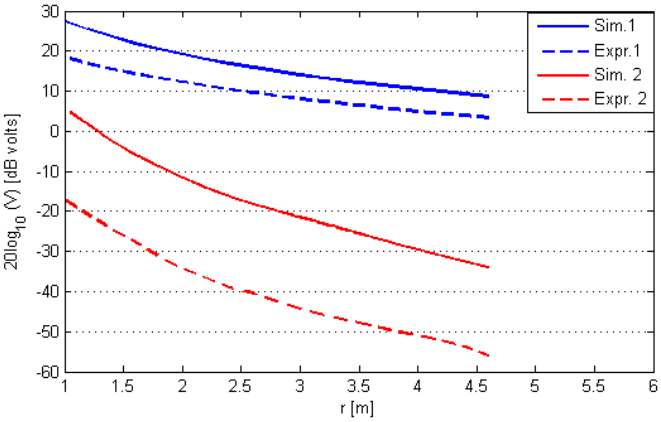

Figure 8.17: Calculated versus measured received voltage v_z for two Tx coils

Calculated and actual voltages are shown in Fig. 8.17. It is obvious that coil 1, as the Tx antenna, is superior to Coil 3 both in simulations and in practice. If we consider a detection threshold of 1 mV, i.e. -60 dBm, the configuration with coils 3 and 2 is useful for no longer than 5 m links. However, the other configuration with coils 1 and 2 looks quite promising at this range. Even better results can be achieved by following coil antenna design criteria of the previous section, more precisely.

Unfortunately, the phenomenon that we observed in Experiment 2, affected this experiment as well. When distance increased, the results became less reproducible. That is why we have not plotted the graph for longer ranges. In summary, the results are sensitive to presence of ferromagnetic objects between transceivers. Effects are not always destructive, and more advanced experiments are needed in order to commercialize this solution.

6 Conclusion

Among the investigated technologies for realizing communication links between pumps of a district heating system, power line communications and magnetic induction are the most viable options. The former has already found many competitive applications. The variety of applications has put off standardization activities. The latter has only been used in very specific domains so far, with high power consumption and extended ranges. It is anticipated in this article that this technology can be used with mild power requirements in underground urban utilities with relatively short range multi-hop networks. The bottom line is to systematically quantify effects of ferromagnetic materials in the vicinity of transceivers. Furthermore, since the operating frequency is very low compared to other RF solutions, available bandwidth is limited. This makes the solution not interesting for broadband applications. Therefore, MI is exclusively useful for sensor networks which require a small amount of bandwidth with guarantees of the availability of the channel as soon as it is needed.

References

[1] WIDE: Decentralized and Wireless Control of Large-Scale Systems, WP2 & WP5. [Online]. Available: http://ist-wide.dii.unisi.it

[2] Neptune project, WP1.1. [Online]. Available: www.shef.ac.uk/neptune

[3] WINES: Wireless Intelligent Networked Systems. [Online]. Available: http://www-civ.eng.cam.ac.uk/geotech_new/WinesInfrastructure/

[4] P^3C: Plug and Play Process Control, WP1. [Online]. Available: http://vbn.aau.dk/en/projects/plug-and-play-process-control-p3c(cdd30235-5d31-4f94-bd17-c29028934471).html

[5] B. Bohm, N. K. Vejen, J. Rasmussen, N. Bidstrup, K. P. Christensen, F. Bruus, and H. Kristjansson, "EFP-2001 Fjernvarmeforsyning af lavenergiomraader," Energistyrelsen, Tech. Rep., March 2004, in Danish.

[6] Danfoss heating book - 8 steps to control of heating systems. [Online]. Available: http://heating.danfoss.com/Content/61078BB5-1E17-4CA2-AC49-8A7CAC2BA687_MNU17424997_SIT54.html

[7] C. D. Persis and C. S. Kallesoe, "Pressure regulation in nonlinear hydraulic networks by positive controls," in *European Control Conference*, Budapest, Hungary, August 2009, pp. 4102–4107.

[8] S. A. Meybodi, P. Pardo, and M. Dohler, "Magneto-inductive communication among pumps in a district heating system," in *the* 9^{th} *International Symposium on Antennas, Propagation and EM Theory*, Guangzhou, China, December 2010, pp. 375–378.

[9] R. Bansal, "Near-field magnetic communication," *IEEE Antennas and Propagation Magazine*, vol. 46, no. 2, pp. 114–115, April 2004.

[10] C. Bunszel, "Magnetic induction: a low-power wireless alternative," *Defense Electronics, Formerly: RFDesign*, vol. 24, no. 11, pp. 78–80, November 2001.

[11] V. Palermo, P. J. Cobler, and N. R. Butler, "Inductive communication system and method," United State Patent 7 254 366 B2, August, 2007.

[12] J. Sojdehei, P. Wrathall, and D. Dinn, "Magneto-inductive (MI) communications," in *Oceans, MTS/IEEE Conference and Exhibition*, Honolulu, HI, USA, November 2001, pp. 513–519.

[13] J. Sojdehei, F. Garcia, and R. Woodall, "Magneto-inductive seismic fence," U.S. Patent 5 969 608, October, 1999.

[14] R. Batcher, "Loop antenna for submarines," *Wireless age*, vol. 7, p. 28, 1920.

[15] D. Reagor and J. Vasquez-Dominguez, "Through-the-earth radio," U.S. Patent 7 149 472, December, 2006.

[16] I. Akyildiz and E. Stuntebeck, "Wireless underground sensor networks: Research challenges," *AdHoc Networks Journal*, vol. 4, no. 6, pp. 669–686, November 2006.

[17] M. Hallikanien, F. Ulaby, M. Dobson, M. El-Rayes, and L. Wu, "Microwave dielectric behavior of wet soil – part I: Empirical models and experimental observations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. GE-23, no. 1, pp. 25–34, January 1985.

[18] M. C. Dobson, F. T. Ulaby, M. T. Hallikainen, and M. A. El-Rayes, "Microwave dielectric behavior of wet soil – part II: Dielectric mixing models," *IEEE Transactions on Geoscience and remote sensing*, vol. GE-23, no. 1, pp. 35–46, January 1985.

[19] I. Akyildiz, Z. Sun, and M. Vuran, "Signal propagation techniques for wireless underground communication networks," *Physical Communication*, vol. 2, no. 3, pp. 167–183, September 2009.

[20] P. Kyosti, J. Meinila, L. Hentila, X. Zhao, T. Jamsa, C. Schneider, M. Narandzic, M. Milojevic, A. Hong, J. Ylitalo, V. Holappa, M. Alatossava, R. Bultitude, Y. de Jong, and T. Rautiainen. (2008, 2) WINNER II channel models, part I, D1.1.2 V1.2. [Online]. Available: http://www.ist-winner.org/WINNER2-Deliverables/D1.1.2.zip

[21] O. Tonguz, A. Xhafa, D. Stancil, A. Cepni, P. Nikitin, and D. Brodtkorb, "A simple path-loss prediction model for HVAC systems," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 4, pp. 1203–1214, July 2004.

[22] M. H. Nayfeh and M. K. Brussel, *Electricity and magnetism*. New York: John Wiley & Sons Inc., 1985.

[23] (2010, August) Wikipedia: Electrical conductivity. [Online]. Available: http://en.wikipedia.org/wiki/Electrical_conductivity

[24] R. Long, P. Cawley, and M. Lowe, "Acoustic wave propagation in buried iron water pipes," in *Proceedings A: Mathematical, Physical & Engineering Sciences*, vol. 459, no. 2039. The Royal Society, November 2003, pp. 2749–2770.

[25] M. Stojanovic, "Underwater acoustic communications: Design considerations on the physical layer," in *Wireless on Demand Network Systems and Services (WONS)*. Garmisch-Partenkirchen: IEEE, January 2008, pp. 1–10.

[26] I. F. Akyildiz, D. Pompili, and T. Melodia, "Underwater acoustic sensor networks: research challenges," *Elsevier Ad Hoc Networks*, vol. 3, no. 3, pp. 257–279, May 2005.

[27] O. Kotlyar, "Universal mud pulse telemetry," U.S. Patent 4 771 408, September, 1988.

[28] I. Wasserman, D. Hahn, H. N. Dang, H. Reckmann, and J. Macpherson, "Mud-pulse telemetry sees step-change improvement with oscillating shear valves," *Oil & Gas Journal*, vol. 106, no. 24, pp. 39–40, June 2008.

[29] H. Sakuma, K. Nakamura, and S. Ueha, "Two-way communication over gas pipeline using multicarrier modulated sound waves with cyclic frequency shifting," *Acoustical Science and Technology, The Acoustical Society of Japan*, vol. 27, no. 4, pp. 225–232, July 2006.

[30] H. Sakuma and J. Fujiwara, "Acoustic communication device and acoustic signal communication method," U.S. Patent 7 027 357 B2, April, 2006.

[31] A. Kondis, "Acoustical wave propagation in buried water filled pipes," Master's thesis, Massachusetts Institute of Technology, February 2005.

[32] G. Kokossalakis, "Acoustic data communication system for in-pipe wireless sensor networks," Ph.D. dissertation, Massachusetts Institute of Technology, February 2006.

[33] Y. Sherif and S. Zahir, "On power-line carrier communication (PLC)," *Microelectronics and Reliability*, vol. 24, no. 4, pp. 781–791, 1984.

[34] R. Arrington, "Load management and automatic meter reading through the use of power line carrier," *Electric Power Systems Research*, vol. 4, no. 2, pp. 85–104, April 1981.

[35] D. Clark, "Powerline communications: finally ready for prime time?" *IEEE Internet Computing*, vol. 2, no. 1, pp. 10–11, January 1998.

[36] S. Galli, A. Scaglione, and Z. Wang, "For the grid and through the grid: The role of power line communications in the smart grid," *Proceedings of the IEEE*, vol. 99, no. 6, pp. 998–1027, June 2011.

[37] R. G. Olsen, "Technical considerations for wideband powerline communication – a summary," in *IEEE Power Engineering Society Summer Meeting*, vol. 3, Chicago, IL, USA, July 2002, pp. 1186–1191.

[38] K. Erickson, A. Miller, E. Stanek, C. Wu, and S. Dunn-Norman, "Pipelines as communication network links," University of Missouri-Rolla, Tech. Rep., March 2005.

[39] C. A. Balanis, *Advanced engineering electromagnetics*. New York: John Wiley & Sons Inc., 1989.

[40] T. Ishida, H. Ando, and M. Fukuhara, "Estimation of complex refractive index of soil particles and its dependence on soil chemical properties," *Elsevier Remote Sensing of Environment*, vol. 38, no. 3, pp. 173–182, December 1991.

[41] E. Locke, "Switchable transceiver antenna," U.S. Patent 6 333 723, December, 2001.

[42] Z. Sun and I. Akyildiz, "Underground wireless communication using magnetic induction," in *In Proc. IEEE ICC*, Dresden, Germany, June 2009.

[43] ——, "Magnetic induction communications for wireless underground sensor networks," *IEEE Transactions on Antenna and Propagation*, vol. 58, no. 7, pp. 2426–2435, July 2010.

[44] K. Kuns, "Calculation of magnetic field inside plasma chamber," UCLA, Tech. Rep., August 2007.

[45] J. C. Maxwell, *A treatise on electricity and magnetism*, ser. Clarendon press. Oxford, 1873, vol. 2.

Paper D

Content-based Clustering in Flooding-based Routing: The case of Decentralized Control Systems

Soroush Afkhami Meybodi, Jan Bendtsen, and Jens Dalsgaard Nielsen

This paper is accepted for presentation in:
The 8th International Conference on Networking and Services, ICNS2012
March 2012, St. Maarten, Netherlands Antilles

Abstract

This paper investigates a problem that is usually studied in communication theory, namely *routing* in wireless networks, but it offers a control oriented solution – particularly for decentralized control systems – by introducing a new routing metric. Routing algorithms in wireless networks have a strong impact on the performance of networked control systems which are built upon them, by imposing latency, jitter, and packet drop out. Here, we have gone one step further from only investigating the effect of such communication constraints, and have directly intervened in the design of the routing algorithm for control systems in order to: 1) realize our preferred network topology and data traffic pattern, and 2) making it feasible to add and remove sensors, actuators, and controllers without having to decommission and/or re-design the system. Moreover, the end-to-end latency and jitter in our system tend to be minimal as a result of robustness of the algorithm to topology modifications. The proposed routing solution combines traditional flooding-based routing scheme with a novel method of clustering nodes based on correlation analysis between existing and emergent sensors and actuators of a control system.

1 Introduction

Chronologically, flooding is the first type of routing solutions that appeared [1]. The name describes well how it works. In pure flooding – that is flooding without any network structure – when a node receives a packet, it checks whether it is the final recipient of the packet or it has received the same packet before. The latter may happen if a packet reaches the same node from different routes. A negative answer to both questions results in retransmission of the packet. Assuming no congestion at the MAC layer, flooding is unquestionably the fastest routing method with the minimum latency which offers a true peer-to-peer (P2P) traffic support. It is not sensitive to modifications of the physical topology because it does not rely on identifying and keeping track of optimal routes. Actually, no topology maintenance is required at all.

All of these advantages come at the expense of a serious drawback. Flooding imposes a heavy overhead, hence is painfully resource consuming. Too many healthy retransmissions happen before a packet is faded from the network. This will not only consume much energy, which is especially valuable for battery-operated nodes, but also causes congestion and increases collision chance at the MAC layer. This diminishes the main benefits of flooding and increases energy consumption both because of the high number of transmissions and the high number of re-transmissions after collisions at the MAC layer happen. That is why pure flooding works well only for small networks with a few number of nodes [1].

Several remedies have been proposed to reduce the number of unnecessary retransmissions in flooding-based routing. Some assume that a limited number of retransmissions are enough to reach the destination and do not propagate the packet any further. Such a number is derived either probabilistically or deterministically by considering the worst case [1]. All of the other methods impose a structure on the network. In the context of flooding-based routing, structuring a network is equivalent to partitioning it into separate or overlapping clusters.

Clustering confines the domain of flooding each packet and can be done by any of the following methods:

- *Coordinates-based Clustering*: Clustering nodes based on their geographical, relative, or virtual coordinates is a simple task provided that the nodes are aware of their geographical location, or can infer their relative or virtual locations. Any kind of distance measure between a node and a cluster center might be used as the membership criterion. If a node is located close enough to the cluster center, it will be a member of that cluster.

- *Metric-based Clustering*: Communication metrics might also be exploited in setting a structure for the network, e.g. by omitting the nodes that have little residual energy or blacklisting the links which are not reliable enough. This category represents a number of popular protocols. Here is how they generally work: At pre-scheduled time intervals, several nodes elect themselves as cluster heads. This could be done by either a pre-defined probability value in each node as in Low-Energy Adaptive Clustering Hierarchy (LEACH) [2], or based on the remaining energy of battery operated nodes as in Hybrid Energy-efficient, Distributed clustering (HEED) [3]. Then the remaining nodes attach to the *nearest* cluster head. "Near" could be interpreted by any kind of distance measure which is derived from a routing metric, e.g. strength of the received signal, expected transmission count (ETX), hop count, etc. [4].

- *Content-based Clustering*: This is a data-centric approach in which the data content of the packet is used to alleviate flooding overhead. Current solutions are either based on tailoring redundant data or aggregating correlated or similar data [5].

In this paper, we are going to introduce another approach towards content-based clustering which is based on the specific characteristics and requirements of the top layer application, i.e. a decentralized control system. Our solution utilizes control oriented metrics to form clusters [6], instead of typically used communication based metrics. Although clustered flooding-based routing is well known, how to create and maintain these clusters makes our routing solution novel.

The remaining of the paper fulfills the above mentioned objectives by coping with the following structure: In Section II, some preliminaries regarding the traffic pattern of our application, which we will call decentralized Wireless Networked Control Systems (WNCS), are stated. They are followed by introducing the assumed networking topology. The main result is given in Section III by proposing a routing algorithm for decentralized WNCS followed by implementation details and step-by-step procedures on cluster formation and maintenance. Some performance related remarks are presented in Section IV. The paper is concluded in Section V.

2 Preliminaries

2.1 Dominant Traffic Pattern

In a WNCS, there are three kinds of nodes: actuators, controllers, and sensors. All of them should have the capability to act both as a data *sink* and as a data *source*, described as follows.

- A Sensor is regularly a data source to send sensory data towards relevant controllers, typically once per control *cycle time* interval, equivalent to the control loop sampling time, e.g. 100 ms.
- A sensor sporadically acts as a data sink to receive configuration data from its associated controllers.
- An actuator is regularly a data sink which receives commands from the associated controller at each control cycle time and implements them.
- An actuator sporadically acts as a data source to report failures.
- Controllers should send and receive data in each control cycle time interval. They gather data from sensors at the beginning of a typical cycle time interval, and send commands to the actuators at the end of the interval. This is exactly what happens in a Programmable Logic Controller (PLC), supplied either with single cable inputs/outputs or aggregated bus communication modules.

Unlike newer routing protocols for WNCSs, which assume a Multi-point to Point (MP2P) traffic with controllers as the sink nodes [7], the above list suggests a P2P traffic pattern. Especially the third item, which has the same importance as the first one and happens at the same frequency, makes WNCSs incompatible with MP2P architecture. Note that, this would not be the case if only monitoring and open loop control were of interest as in [7].

2.2 Network Topology

Fig. 9.1 shows the logical network topology of a simple decentralized WNCS. Arrows represent data direction in its regular functioning mode, but information flow in the reverse direction is also required as explained earlier in II.*A*.

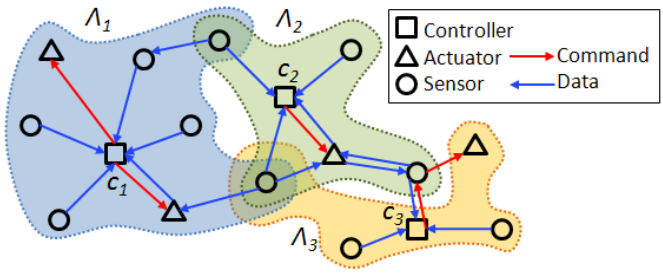

Figure 9.1: Logical network topology of a simple decentralized wireless networked control system

Fig. 9.1 depicts the key assumptions that we have considered in topology design, described in the following:

1. There could be a large number of controllers ($C_i, i = 1, ..., n$).

2. Each controller and its associated sensors and actuators form a set, called a *cluster* henceforth. Clusters are identified by their unique tag ($\Lambda_i, i = 1...n$).

3. There are as many clusters as controller nodes which are called *cluster heads*.

4. Each packet contains a cluster *association* field. In general, a packet might be tied to one or more clusters. It is also possible that a packet is not associated with any cluster.

5. A sensor might be a *member* of multiple clusters, meaning that its generated data could be associated to more than one cluster head. In other words, a packet that is generated at a sensor node, might be reported to more than one controller.

6. An actuator could be a member of at most one cluster, meaning that it may not receive commands from more than one controller.

7. Data packets to/from members of a cluster should be sent from/to the cluster head. In other words, a controller node is either the source or the final destination in every transmission path.

8. Relaying data packets associated to the cluster (Λ_j) could be done via all of the nodes in the entire network, irrespective of their membership status in Λ_j, provided that the relayed packet has enough remaining *relay credit*.

Relay credit can be defined in terms of any scalar node or link routing metric, e.g. hop count as a node metric or expected transmission count (ETX) as an accumulative link metric [4].

The last assumption clarifies that cluster membership is not necessary when relaying a packet. The purpose of assigning relay credit to a packet is to preserve connectivity in a cluster while constraining the number of retransmissions that might occur to a packet among non-member nodes. Each data packet is given an initial relay credit besides cluster membership tags, when generated. While it is roaming inside its own cluster, it does not spend any relay credit. However, when it is being relayed among non-member nodes, the initial relay credit is decreased at each non-member node until the remaining credit is not enough for more retransmissions among non-member nodes. Fig. 9.2 illustrates an example.

In Fig. 9.2, the dark area shows the members of a cluster Λ, and the lighter surrounding area shows the maximum penetration depth of the packets of Λ if hop-count ≤ 2 is considered as the relay credit criterion. It helps preserving connectivity of Λ members, when there is no direct physical link between them. This is shown in Fig. 9.1 too, where a sensor node which is a member of Λ_3 connects to its own cluster via an actuator and a sensor of Λ_2.

3 The Routing Protocol

3.1 Control Oriented Clustering of Nodes

To propose a method to form clusters based on requirements of the control system, we rely on the results of [6]. It offers three stochastic correlation-based measures that indicate

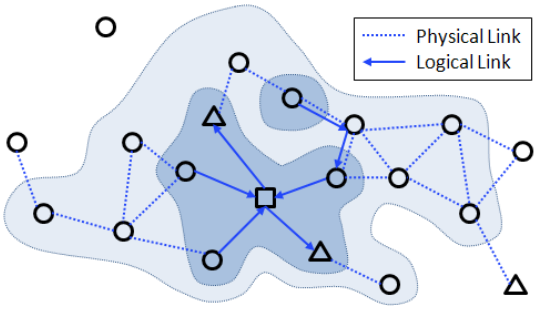

Figure 9.2: Preserving connectivity of cluster members by a surrounding cloud whose thickness is adjusted by *relay credit*, here defined as hop-count ≤ 2

usefulness of incorporating a new sensor/actuator in a present system model. We will use these scalar measures as application-based routing metrics in order to extend the concept of *distance*. Consequently, membership in a specific cluster Λ_j is granted to a node if that node is closer to the cluster head C_j than a specific threshold value d_j.

3.1.1 Addition of a new sensor

For a newly added sensor node, the following two complementary measures are proposed.

$$d^2_{U_p,y_a} = \frac{E[y_a(k) - \hat{y}_a(k|U_p^{k-1})]^2}{E[y_a(k) - E(y_a(k))]^2} \tag{9.1}$$

$$d^2_{U_pY_pY_a,y_a} = \frac{E[y_a(k) - \hat{y}_a(k|U_p^{k-1}, Y_p^{k-1}, Y_a^{k-1})]^2}{E[y_a(k) - E(y_a(k))]^2} \tag{9.2}$$

in which d^2 represents the correlation based distance and varies between 0 and 1. Subscripts $(.)_p$ and $(.)_a$ refer to present model and added device, respectively. $y(k)$ and $u(k)$ mean individual samples of a sensor's data and an actuator's command at time k, while Y^k and U^k indicate the set of all samples from the beginning up to and including time k. $E(.)$ stands for expected value operator over a finite number of data samples, N, which is pre-defined in the sensor node. Superscript $\hat{(.)}$ stands for the least-squares estimation based on the available model.

With respect to the above mentioned definitions, interpretation of (9.1) and (9.2) is given in the following paragraph, assuming that: 1) the model of the present system is discrete-time linear time-invariant and 2) the new sensor provides sufficiently exciting data, and 3) a consistent un-biased least-squares estimation is given when $N \to \infty$.

The denominator in (9.1) and (9.2) is the variance of the data gathered by the new sensor, i.e. y_a. The numerator in (9.1) indicates how predictable the current $y_a(k)$ is if the commands of all actuators are known in the previous samples. If, according to the present model, none of the $k-1$ samples of all of the actuators have any tangible effect on the k^{th} sample of y_a, the following equation holds true.

$$E[y_a(k)|U_p^{k-1}] = E(y_a(k)) \tag{9.3}$$

Furthermore, if an unbiased estimation is assumed, we have $\hat{y}_a(k|U_p^{k-1}) = E[y_a(k)|U_p^{k-1}]$ which in combination with (9.3) results in the following expression:

$$\hat{y}_a(k|U_p^{k-1}) = E(y_a(k)) \tag{9.4}$$

Equation (9.4) means that the conditional least squares estimation of y_a is equal to its actual expected value. Therefore, the present model is good enough and the new measurement does not add any value to it. In this situation, $d^2_{U_p,y_a} = 1$, which should be read as: the new node is too far from the cluster head and cannot become a member, that is it is irrelevant to the control loop in question.

Equation (9.1) measures how much the additional sensor is affected by the present actuators in open loop. Nevertheless, this measure only reveals linear correlation. To look for nonlinear correlations, (9.1) should be modified according to the specific nonlinearity we are looking for. This is the easy step, but the difficult part is to perform nonlinear online incremental system identification to find $\hat{y}_a$. We do not consider this case in this paper.

The numerator in (9.2) measures how much the additional output could be controlled by the present actuators in closed loop. The interpretation is similar to (9.1), but this time the data from all of the sensors, including the new one, are also used in the least squares estimation, hence making it a more computationally intensive problem. Either (9.1) or (9.2) could be used in a given setting. Exploiting (9.1) is recommended in cases where y_a cannot be controlled independently of y_p [6].

3.1.2 Addition of a new actuator

When a new actuator is added, the following measure is proposed.

$$d^2_{U_a y_p|U_p Y_p} = \frac{E[y_p(k) - \hat{y}_p(k|U_p^{k-1}, Y_p^{k-1}, U_a^{k-1})]^2}{E[y_p(k) - \hat{y}_p(k|U_p^{k-1}, Y_p^{k-1})]^2} \tag{9.5}$$

Equation (9.5) measure how much influence the additional actuator has on the present sensors in closed loop. If U_a^{k-1} does not have any effect on improving prediction of $\hat{y}_p$, then the prediction errors in the numerator and denominator will look alike and d will get its maximum value ≈ 1 On the other hand, if U_a^{k-1} is useful such that the prediction error in numerator is much less than that in denominator, then we have: $d \to 0$. Equation (9.5) should be interpreted similar to the previous measures with similar concerns. The same assumptions hold for a consistent estimation of $\hat{y}_p$. In practice, to provide a sufficiently exciting control signal, the actuator has to be driven by an external signal.
Remark 1: Latency in the communication network has a considerable impact on all of the introduced measures. But at the same time, it influences control performance too. Therefore, it is reasonable to consider this effect on evaluating usefulness of adding new sensors and actuators.

3.2 Network Layer Packet Format

To illustrate details of the protocol, the packet format shown in Fig. 9.3, is chosen for the Network layer.

The first field in the packet is a set of bit registers. The first bit flag determines whether the packet is sensor or actuator related. A sensor packet is generated in one of the sensors

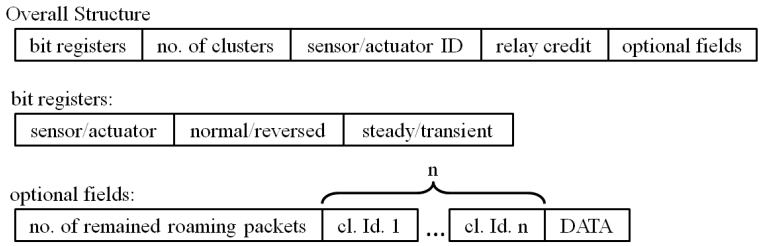

Figure 9.3: Structure of data packets at Network layer

and should be routed to one or more controller nodes. An actuator packet is generated in a controller node and should be routed to an actuator node. The second bit indicates the direction of data. For sensor packets, "normal" means "from sensor to controller" and "reverse" means the other way around. The converse is true for actuator packets. The third bit register shows if the packet is sent in *steady* or *transient* operating mode. In transient mode, the packet contains an additional field, namely *number of remained roaming packets*, which is listed among optional fields. The difference between these two modes and the function of roaming packets is explained later in Section III.C.

The second field stores the number of associated clusters. "Zero" in this field means that the packet is not associated with any cluster. If the packet is linked to $n > 0$ clusters, n additional fields are included in the packet, each of which containing the address of one of the associated clusters.

The next field is *Sensor/Actuator ID*. In sensor packets, it contains address of the sensor node that has generated the packet. In actuator packets, it contains address of the actuator node that the packet is destined to. Assignment of unique addresses to all of the nodes in the entire network is a prerequisite to our routing solution. The same demanding requirement exists in emerging standards, e.g. IETF ROLL, by incorporating IPv6 as a worldwide addressing standard [7].

The packet also contains a *relay credit* ≥ 0 which is explained in details, earlier in Section II.B.

The last optional field is DATA which actually contains the application layer packet.

3.3 Cluster Formation

Here, we give a high level description of cluster formation. Assume that the controller nodes ($C_i, i = 1, ..., n$) are deployed as cluster heads. In a realistic scenario, each cluster head has a built-in model of the subsystem it is supposed to control. All of the initially deployed sensor and actuator nodes are already bound to their controllers. In other words, in the network setup phase, all of the nodes are aware of their cluster membership. As a result, sensor nodes immediately start to function in their normal operating mode. See Fig. 9.4.

Actuator nodes should start working in a safe mode and wait until they receive commands from the cluster head, i.e. the controller unit. The cluster head starts sending commands to the actuator as soon as it can devise the commands based on received sensor packets and the pre-programmed plant model. Command carrying packets are illustrated

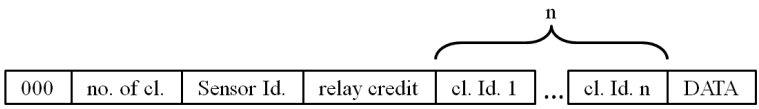

Figure 9.4: Structure of sensor packets at steady operation

in Fig. 9.5.

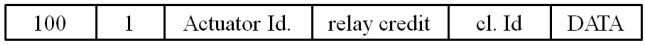

Figure 9.5: Structure of actuator packets at steady operation

In both above cases, the packets flood their pertinent clusters. Moreover, they propagate among the nodes of neighboring clusters into a certain depth defined by their relay credit.

Later on, when a new sensor pops up, it does not initially belong to any cluster and it is in transient operating mode. Thus, it publishes data in *roaming packets* as shown in Fig. 9.6.

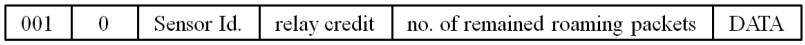

Figure 9.6: Structure of packets generated by a sensor when it is just turned on

When a roaming packet arrives at a neighboring node that is operating in *Steady* mode, it inherits all of the cluster tags of that node – meaning that the $cl.Id.$ fields of the packet are refreshed. This action is performed only if *relay credit* > 0. If either the neighboring node is in *Transient* mode or *relay credit* $= 0$, cluster tags of the packet remain unchanged.

In steady mode, embedding a cluster tag into a packet gives it the right to freely flood in that cluster without spending relay credit, but it is not the case in transient mode in which relay credit is constantly spent for every transmission. Therefore, embedding cluster tags into roaming packets does not give them a free pass. On the other hand, running out of relay credit is not the stop criterion when retransmitting a packet in transient mode. It just kills its ability to inherit new cluster tags. At the end, a roaming packet floods into the clusters that it managed to enter before running out of relay credit.

Example 1: Fig. 9.7 depicts an example when a new sensor is placed amongst nodes of Λ_1. However, some of its roaming packets could also reach the borders of Λ_2 before consuming all of their relay credit. Thus, presence of the new sensor is advertised through the union of nodes of Λ_1, Λ_2, and in the *relay credit* > 0 zone. Based on the above setting, C_1 and C_2 start calculating (9.1), or (9.2), or both. Note that, C_3 might also receive the roaming packets of the new sensor if it is placed in $\Lambda_3 \bigcap \Lambda_1$ or $\Lambda_3 \bigcap \Lambda_2$, but it will not calculate (9.1) or (9.2) because Λ_3 is not listed among clusters in the packets.

Fig. 9.8 shows how the correlation-based distance measures (1) or (2) might develop over time in C_1 and C_2. It is assumed that the new sensor generates 1200 roaming packets.

After collecting sufficient samples at the cluster heads, each C_i decides whether the new sensor should become a member of their cluster or not. In Fig. 9.8, C_2 concludes that

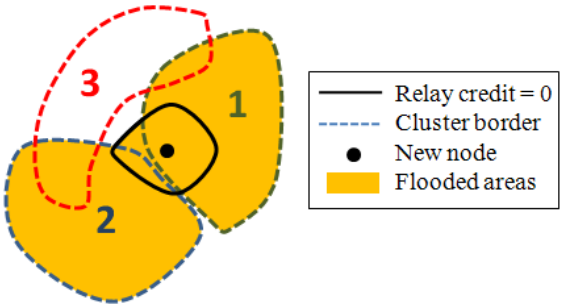

Figure 9.7: Effect of *relay credit* when a new node is joined

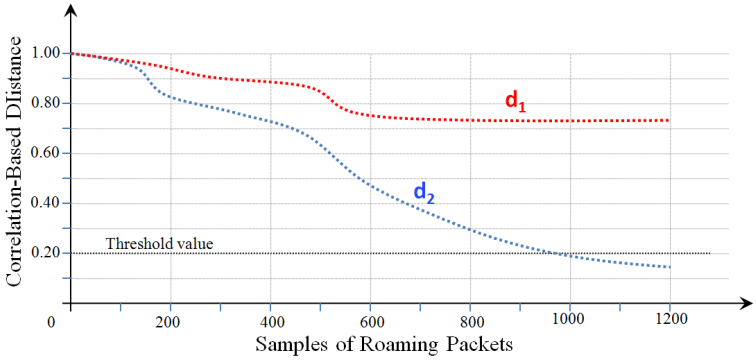

Figure 9.8: Development of usefulness measure (1) in C_1 and C_2 during identification process

the new sensor is relevant well before the roaming packets are discontinued, but C_1 finds the new sensor irrelevant to the control performance of its internal model. C_2 continues the joining process by sending a *join* request to the new sensor. The data packet which contains the join request is in the following form, shown in Fig. 9.9.

010	1	Sensor Id.	relay credit	cl. Id. 1

Figure 9.9: Join request from cluster head to a sensor

Note that, this packet is in ”steady operating mode”, which means no relay credit is deducted unless it leaves the source cluster, Λ_2 in our example. This mechanism is useful when the new node is only reachable via nodes of other clusters.

After sending all roaming packets, equal to 1200 in our example, the new sensor applies received join requests. Join requests arrive at the sensor node asynchronously. Therefore, the node should continually accept join requests, at least until a pre-defined time. If no join request is received, the sensor starts another round of generating roaming packets.

If a group of sensors are deployed simultaneously, the ones which have other cluster

members in their vicinity will find their clusters earlier. The others that do not have any neighbor operating in steady mode, may not inherit cluster tags, hence their flooding domain is limited to the *relay credit* > 0 zone.

The above procedure is slightly different for a new actuator. Each newly turned on actuator applies a pre-specified control sequence for the purpose of sufficiently exciting the plant and creating measureable outcomes. Simultaneously, it publishes roaming packets as shown in Fig. 9.10, which contain current value of the actuator output.

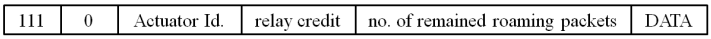

Figure 9.10: Packets generated by an actuator when it is just turned on

The cluster heads which receive these packets, use the DATA field in evaluating (9.5) similar to what was shown in Fig. 9.8. When the roaming packets are discontinued – meaning that the actuator is waiting for the decision – each cluster head returns the calculated *usefulness measure* to the actuator by packets shown in Fig. 9.11.

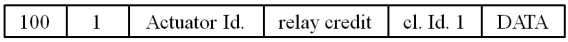

Figure 9.11: Packets that return usefulness of utilizing an actuator in a cluster

The actuator waits for a certain time to receive evaluation results from all involved cluster heads. Then it compares the received usefulness values, which are embedded in DATA fields, and selects the cluster that has returned the highest value. If at least one evaluation result is received, the actuator chooses its own cluster and sends a join request to that cluster as illustrated in Fig. 9.12. If no evaluated usefulness measure is received, the above procedure starts from the beginning.

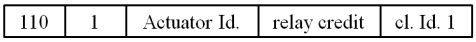

Figure 9.12: Join request from an actuator to a cluster head

Note that, when a new sensor is added, it receives individual join requests from controllers. But when a new actuator is added, it sends the join request to a single controller.

4 Performance Related Remarks

Remark 1: Unlike other clustered flooding-based routing protocols whose acceptable performance depend heavily on the optimal choice of the number of clusters and the thoughtful selection of cluster heads [2, 3], these parameters are pre-defined in our protocol because all of the controller units ($C_j, j = 1...n$) and only the controller units are cluster heads. Moreover, the controller nodes are fixed in the whole lifetime of the network.

Remark 2: Another issue is the influence of lower layer protocols on performance of the routing protocol. Unlike [2], that has utilized a TDMA-based MAC in sake of energy-efficient collision-free transmissions, the MAC layer in our system cannot accommodate a deterministic reservation-based protocol. It is mainly due to the constrained coverage

range of nodes which does not guarantee existence of direct links between cluster heads and every member of the cluster to schedule a frame-based MAC. After all, it is a prerequisite for framed MACs that all of nodes can be accessed from a single base station for scheduling purposes. Otherwise, many time frames must be kept unused and reserved for future extension, as in Time Synchronized Mesh Protocol (TSMP) [8].

Our routing solution may be built either on a contention-based or a preamble-sampling MAC which are inferior to deterministic MACs in terms of energy efficiency and end-to-end latency if data transmission among nodes is frequent.

Therefore, the main benefit of our routing solution is to pick the members of each cluster so prudently such that it results in the minimum number of nodes in a cluster, and more efficient flooding in clusters.

5 Conclusion

In this paper, we have proposed a method to form clusters of nodes to be utilized by a flooding-based routing algorithm in decentralized wireless networked control systems. Our cluster formation method is data-centric and originates from the requirements of the control application. It makes use of model-based correlation estimation between new nodes and the existing model of the system. Furthermore, we have proposed to exploit non-member nodes in providing connectivity among nodes of an individual cluster. To this end, we have suggested to use an arbitrary routing metric, e.g. hop count. Operation of the proposed clustering mechanism is described in details.

References

[1] T. Watteyne, A. Molinaro, M. G. Richichi, and M. Dohler, "From MANET to IETF ROLL standardization: A paradigm shift in WSN routing protocols," *IEEE Communications Surveys & Tutorials*, vol. 13, no. 4, pp. 688–707, Fourth Quarter 2011.

[2] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 660–670, October 2002.

[3] O. Younis and S. Fahmy, "Heed: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Transactions on Mobile Computing*, vol. 3, no. 4, pp. 366–379, October-December 2004.

[4] IETF ROLL: routing metrics used for path calculation in low power and lossy networks. [Online]. Available: http://tools.ietf.org/wg/roll/draft-ietf-roll-routing-metrics/

[5] J. Kulik, W. Heinzelman, and H. Balakrishnan, "Negotiation-based protocols for disseminating information in wireless sensor networks," *Wireless Networks*, vol. 8, no. 2, pp. 169–185, March-May 2002.

[6] T. Knudsen, J. Bendtsen, and K. Trangbaek, "Awareness and its use in incremental data driven modelling for Plug and Play Process Control," *European Journal of Control*, vol. 18, no. 1, pp. xx–yy, February 2012.

[7] IETF ROLL: IPv6 Routing Protocol, draft standard. [Online]. Available: http://tools.ietf.org/wg/roll/draft-ietf-roll-rpl/

[8] K. S. J. Pister and L. Doherty, "TSMP: time synchronized mesh protocol," in *Proceedings of the IASTED International Symposium on Distributed Sensor Networks*, November 2008.

Paper E

Wireless Plug and Play Control Systems: Medium Access Control and Networking Protocols

Soroush Afkhami Meybodi, Jan Bendtsen, Jens Dalsgaard Nielsen, and Torben Knudsen

This paper is submitted in:
European Journal of Control

The layout has been revised

Abstract

Plug and play control systems are gradually finding their place as a new class of control systems that offer high flexibility in addition and removal of sensors and actuators. This paper focuses on protocols of the communication framework for wireless plug and play control systems. It is started by introducing the main challenges in wireless control system design with stringent requirements on link reliability and end-to-end latency. Thereafter, several selected medium access control protocols are reviewed and a few are chosen as the recommended options based on some factors including sensitivity to latency and jitter. A similar procedure is followed for choosing the most viable routing solutions. Assumptions on the networking topology turned out to have major importance in finding the best options in the networking layer. The main contribution of the paper, however, is introduction of a new routing metric that is truly application-based and is completely determined by the control system. Such a routing metric makes a direct link between the top level application layer and the networking layer by implementing the networking topology in accordance with the control topology. To illustrate usefulness of the proposed routing metric, it is shown in details and via simulations, how it can be exploited in clustering the nodes of a wireless networked control system with respect to their relevance to individual control loops.

Keywords: *Plug & play control, Topology, Flooding-based routing, IEEE802.15.4e, IEEE802.11p*

1 Introduction

The term *Plug and Play* (PnP) is well established when it comes to computer accessories. It basically refers to a device with specific known functionalities that could be utilized as soon as it is attached to the computer, with the least required effort to install and run it. The main idea behind PnP control is to extend the concept of PnP from computer accessories to the building blocks of a control system which are sensors, actuators and sub-systems with more general functionalities.

The variety of applications of control system elements makes their employment not as straightforward as computer accessories. For instance, when a USB flash memory is attached to a computer, the operating system makes it ready to be read/written as a temporary mass storage device. But, in a similar scenario, what should a control system do if a new temperature sensor pops up and starts reporting temperature of a vessel in a large process to the controller unit?

Normally, designing control logics and algorithms is the last step in the system setup, after a model is developed for the existing system with chosen sensors and actuators. Making sensors and actuators plug and play, requires the controller to be able to identify and cope with changes in the system model and sensor-actuator pairing. Building a PnP control system involves several tasks, namely:

1. Automatically making communication links between added devices and the existing system by a self-organizing self-healing communication network

2. Identifying types/locations of instruments that have joined the system, and finding what kind of measurements/actuations they offer.

3. Automatically updating the system model which is stored in the controller(s).
4. Automatically updating control algorithm(s) to maintain/improve control performance.

Several works have recently appeared with focus on the above tasks and various aspects of PnP control systems [1, 2, 3, 4, 5]. This paper looks into the infrastructure of wireless PnP control systems, i.e. reconfigurable scalable wireless communication networks which can tolerate removal (deliberately or by faults) and addition of devices without interruption, and interference with other existing systems. An illustrative example of a wireless PnP control system block diagram and a possible networking topology are shown in Fig. 10.1 and Fig. 10.2, respectively. In general, a block diagram may correspond to many network topologies, and a network topology may correspond to many block diagrams.

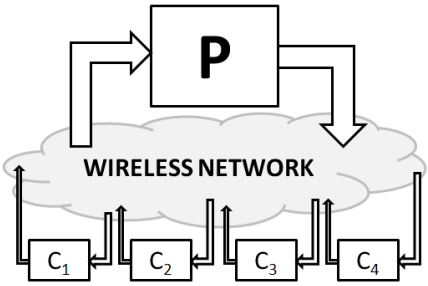

Figure 10.1: Block diagram of a wireless PnP control system with four multi-input single-output controllers $C_1 ... C_4$

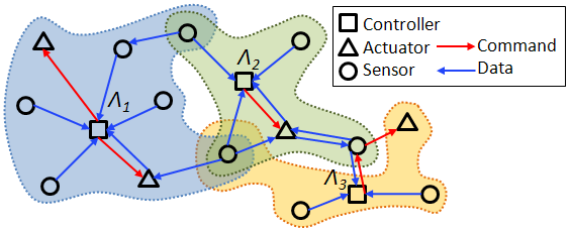

Figure 10.2: Networking topology of the wireless PnP control system in Fig. 10.1 implemented with three controller units $\Lambda_1 ... \Lambda_3$

Although there is a growing interest in industrial wireless networks, wireless turned out not to be a viable replacement for wire. Gradually, it was accepted that a different scope and application domain with milder timing requirements should be assigned to industrial wireless networks [6]. *Equipment monitoring*, *asset tracking*, and *open loop control* are the representative examples. This strategy considers wireless only as a complement to wireline networks, in order to extend their functionality. In spite of this fact, this paper looks into applications with tougher timing requirements, that is *closed loop control* in Wireless PnP control systems.

Industrial wireless networks lack of performance compared to wireline networks is not due to immaturity of the technology. It is because of the extra problems that should be dealt with in a wireless framework. These problems are inherent and always do exist. How we manage to deal with them defines how close the performance of industrial wireless networks could get to their wireline counterparts. Two of these particular challenges are: 1) stochastic nature of the quality of wireless links which affects reliability; and 2) additional end-to-end latency due to the limited coverage range of wireless nodes in a multi-hop topology.

1.1 Lack of Reliability

Reliability is of utmost importance in industrial networks. It is generally defined as the fraction of the sum of healthy received packets at all nodes over the sum of transmitted packets from all nodes. There is no failed transmission in a 100% reliable network. Lack of reliability in wireline networks is mainly due to collisions at the *medium access control (MAC)* layer. In wireless networks, however, there is an additional cause for it which is variation in the quality of wireless links. To enhance reliability, we have selected and compared several MAC protocols that can alleviate the effect of uncertainty in quality of wireless links, and hence offer deterministic behavior.

1.2 Additional End-to-End (e2e) Latency

Besides the inevitably influential parameters on e2e latency, including the *physical layer (PHY)* parameters and the MAC protocol, there exists another source of e2e latency in multi-hop wireless networks which is not always deterministic and introduces variations in latency, i.e. *jitter*. Such additional e2e latency occurs when a packet needs to be relayed via a number of intermediate nodes to arrive at its final destination. In order to minimize the e2e latency and routing overhead in wireless multi-hop PnP control systems, we have proposed a new application-based routing metric which makes a direct link between the application layer and the networking layer by implementing the networking topology in accordance with the control topology.

In order to achieve the above goals, the paper continues in Section II by reviewing a few promising MAC protocols that are competing to attain acceptance and dominance in industrial applications. We will pick up the most viable options in favor of wireless PnP control systems. Section III tackles the routing problem and hence, reviews relevant *routing* protocols in multi-hop wireless networks. Introducing the problem setup of a wireless PnP control system, its predominant data traffic pattern and networking topology precede the main contribution of the paper in Section IV, which is proposing a new routing metric and its application in flooding-based routing. Thereafter, Section V merges the results from previous sections in a simulation study where the presented routing solution is evaluated in connection with the MAC protocols that are recommended earlier in Section II. Finally, Section VI concludes the paper.

2 MAC protocols of Wireless PnP Control Systems

2.1 Token-based MAC Protocols

Token passing is originally designed for wireline networks. It is the MAC protocol for industrial wireline networks such as Profibus FMS and CAN bus [7]. Thus, it would be very interesting if it is also applicable to industrial wireless networks.

In Token passing MAC protocols, a single token is transferred in a ring among the nodes that might need to transmit data. Whoever possesses the token has the right to transmit. Token passing method offers a deterministic behavior in controlling channel access. It guarantees Quality of Services (QoS) in terms of bounded latency and reserved bandwidth by constraining the maximum number of nodes that can join a ring and the longest time interval that a node may possess the token before passing it to the next node. Token passing is a distributed mechanism and does not require a base station to access all of the nodes in order to synchronize or schedule them. Therefore, it is a potential candidate for wireless networks with limited coverage range. However, it has been shown in [8] and [9] that the performance of token passing MAC protocols over lossy wireless links may easily fall below the acceptable level. It is due to high sensitivity of the protocol to reliable transmission of *token transfer* messages. Temporary loss of the link that should pass the token to the successor node will result in temporary exclusion of the successor from the token ring and prevents it from transmitting any data irrespective of its importance. It is called the *ring stability problem* [8]. Therefore, it is not suitable to employ the MAC protocol of e.g. wired Profibus, in order to create the wireless version of this popular field bus.

Several researchers have tried to offer a wireless token passing protocol. A remarkable result is the Wireless Token Ring Protocol (WTRP) [10]. WTRP handles automatic topology reconfiguration when nodes join and leave rings. The main contribution of WTRP is introducing several other control messages named after different kinds of tokens. Some of them deal with ring maintenance while the others handle automatic reconfiguration and scalability. However, the main concern still exists. WTRP gives proof for ring stability assuming no link failure. Although a ring recovery mechanism is introduced, there is no evaluation of ring stability when the links are lossy. Thus, WTRP does not guarantee ring stability in presence of lossy links [10].

2.2 Reservation-based MAC: Wireless Personal Area Networks

Among the originally designed solutions for wireless networks, Wireless Personal Area Networks (WPANs) introduce an increasingly popular category. WPAN is the subject of the IEEE802.15.4 PHY and MAC layer specifications for local and metropolitan area networks. It focuses on multihop wireless networks which consist mainly of battery-operated resource-limited nodes. WPANs have found many applications including home automation e.g. in Zigbee, industrial automation e.g. in WirelessHart and ISA100.11a, large scale environmental monitoring, asset tracking, etc.

WPANs make use of *Reservation-based* protocols in the MAC layer [11]. Reservation-based MAC protocols provide a time schedule for every node, such that each one is aware when to access the channel and for how long keep control of the channel. The scheduling node in reservation-based MACs requires knowing the complete and up-to-date topology

of the entire network. Moreover, providing some sort of synchronization is essential to establish a common sense of time among all nodes. These demanding requirements potentially result in ideal performance and optimal use of resources. On the other hand, lack of scalability and reconfigurability are the main problems for reservation-based MACs.

The IEEE802.15.4 is amongst the new reservation-based protocols that use a hierarchical network topology which provides flexibility and scalability to the network. For instance, there could be different types of nodes with different levels of functionality. Some *reduced function* nodes may only contribute as leaves in the star topology while the *full function* nodes could make a mesh network amongst themselves and also act as the gateway for reduced function nodes. In such a network, it is possible to assign different MAC parameters for different types of nodes.

Since ratification of the latest version of the IEEE802.15.4 in 2006, several amendments have appeared to enhance its functionality in various directions among which activities in the *4e* working group deserves more scrutinization. IEEE802.15.4e amends MAC enhancements to 802.15.4 for *industrial applications*. Its main purpose is to add two features of *high reliability* and *low latency* which are the most serious concerns in industry. 802.15.4e is in draft status yet, but some of its basic contents are mentioned in the sequel [12].

Application domains in 802.15.4e are divided into two groups based on their latency requirement, namely the latency tolerant Process Automation (PA) and Commercial Applications (CM), and the latency sensitive Low Latency (LL) networks. The key aspect in PA, CM and LL cases is *determinism*; with the lowest possible deterministic latency for all LL, a tunable deterministic latency for PA and CM, and the highest feasible reliability, i.e. deterministic link quality for LL, PA and CM applications.

2.2.1 Deterministic Latency in the IEEE802.15.4e

In LL networks, a device could be either a sensor with merely uplink traffic or an actuator with bidirectional up/downlink traffic (Fig. 10.3). The gateway schedules TDMA access – one fixed time slot per node – for all of the nodes to create a superframe of timeslots. Superframe duration depends on the number of devices. It can be in order of 10 ms for a typical network of 20 devices. Superframes are synchronized by a beacon message transmitted frequently from the gateway.

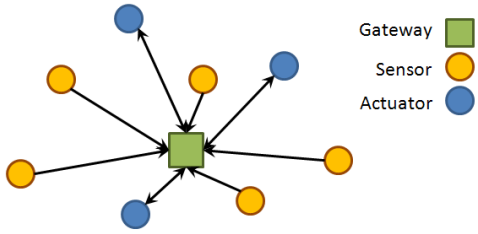

Figure 10.3: Network topology for low latency application

Special MAC frames with 1-octet MAC header are proposed to reduce MAC overhead, hence facilitating smaller PHY packets that fit into short timeslots and make superframes shorter.

Clearly, the use of multiple transceivers on different channels accommodates more devices in a superframe with specified length, but since there could be many LL networks in the proximity of each other, utilizing multiple channels is not allowed in sake of providing coexistence for LL networks.

For latency tolerant PA applications, an extended version of Time Synchronized Mesh Protocol (TSMP) is used [13]. TSMP is compliant with 802.15.4 and is the main ingredient of WirelessHART and ISA100.11a. Facing the problem of limited coverage range of nodes, while requiring a deterministic cyclic performance, forced TSMP inventors to consider superframes with fixed duration and lots of unused time slots which were reserved for accommodating the pre-defined maximum number of nodes that might join the network. This conservative approach makes the performance of TSMP independent of the number of nodes by assuming the worst case. The length of the superframe defines e2e latency and is determined by the maximum number of nodes and an assumed physical distribution of nodes. If the nodes are placed far away from each other, more retransmissions are needed among multiple hops, and hence a longer superframe is required.

2.2.2 Towards Reliable Wireless Links in 802.15.4e

In LL networks that assumingly take up small areas, reliability is provided by prudently selecting a single channel and constraining presence of sources of interference in the proximity. For PA networks, an improved version of TSMP, known as Time Slotted Channel Hopping (TSCH) [14], seeks the same objective.

In 802.15.4e, a link is defined by a time slot and a channel frequency offset. 802.15.4e makes use of 16 narrow-band channels in the interval 2.400-2.500 GHz. Unlike *channel agility*, also called *channel adaptation*, in which the channel changes only after a problem occurs in the current channel, TSCH suggests a pre-planned channel hopping. In TSCH, an algorithm switches among different channels at each timeslot, with the ability to black-list some channels in sake of providing coexistence with other networks. Channel hopping and channel adaptation constitute *channel diversity* for 802.15.4e networks, which considerably increases probability of having interference-free wireless links.

2.3 Contention-based MAC: Vehicular Ad-hoc Networks

Vehicular Ad-hoc Networks (VANETs) introduce another active area in standardization of wireless protocols for specific applications. VANET, as introduced in the IEEE802.11p standard, refers to wireless networks whose nodes may dissipate energy in order to keep up with the quickly changing topology and the tight e2e latency requirements of vehicular wireless networks.

VANET employs a *Contention-based* MAC solution. This family of MAC protocols in contrast to reservation-based MACs is known for its simplicity which is attained by sacrificing performance. Nodes with contention-based MAC may have common active periods and are prone to collision occurrence at the PHY. In comparison with reservation-based MACs, performance of contention based protocols is always inferior. However, the difference depends on traffic congestion. In unsaturated low load networks the measures of performance, e.g. throughput or reliability, are very close for both categories. As traffic increases, performance of reservation based protocols remains constant until traffic saturation, but performance of contention based protocols begins to degrade and ends up

in a jammed situation, well before the saturation point. That happens when no healthy transmission is practically possible and the retransmission requests are accumulated to make a virtual saturation.

While the IEEE802.11 defines PHY and MAC layer specifications for wireless LAN networks in general, the amendment of 802.11p for VANETs specifies PHY and channel access mechanisms which suit vehicular environments. These include vehicle to roadside and inter vehicle communications with stringent requirements on latency which are the same as wireless PnP control systems

The IEEE802.11p enjoys seven channels at the PHY, designated to be used for different applications ranging from spreading time critical safety messages, e.g. vehicle collision warnings, to distributing supportive data, e.g. updated traffic information or even local commercial advertisements. Hence, all packets are marked with their priority, namely Access Category (AC), to be able to gain access to the appropriate channel. One of the channels is solely designed as the control channel and the other six could be used for data handling.

At the MAC layer, 802.11p has adopted the IEEE802.11e. 802.11e amends MAC enhancements to 802.11 in order to improve QoS. In general, 802.11e has two operating modes. One is based on polling nodes from a central station, hence is a contention-free mode and assigns guaranteed time slots to specific nodes. It is called *hybrid coordination function controlled channel access (HCCA)*. The other, which is the main operating regime in VANET, relies on distributed CSMA/CA and is called *enhanced distributed coordination access (EDCA)* [15].

The basic idea behind EDCA mechanism is to prioritize packets based on their data-content. It defines individual sets of channel access parameters such that high priority packets have a higher probabilistic chance to win the contention. The mentioned parameters are: 1) transmission delay after sensing the channel idle, and 2) the back-off time after unsuccessful transmissions. For example, when a high priority packet shall be transmitted, both of the parameters get smaller values. In summary, the EDCA mechanism in VANETs provides stochastic prioritization with adjustable parameters.

In automation industry, the focus has been on WPAN so far, but it is understood that latency sensitive applications and closed loop control are far from being feasible in the current practice where the main concern is energy consumption. At the same time, high reconfigurability and tight e2e latency requirements of PnP control systems, bring VANETs into consideration. To further develop this initiative, the key properties of WPANs and VANETs are summarized in the first two columns of Tab. 10.1, followed by the properties of an ideal network that hosts a PnP control system. We call it Controller Ad hoc Network (CANET) henceforth.

The similarities between VANETs and CANETs in Tab. 10.1 suggest to eliminate the constraint on power consumption in CANET in order to provide a chance to satisfy requirements of latency sensitive applications. This means that the nodes will not be truly wireless since they have to keep their power cords. However, getting rid of signal cables is an advantage by itself. It is reasonable, at least for *actuators* including motors, pumps, and similar devices to always run on power cables and have their own dedicated electrical infrastructure in factories and industrial plants.

Table 10.1: Comparison of key properties in designing MAC protocols for wireless PnP control systems

Network Type	VANET	WPAN	CANET
Constrained Power	No	Yes	??
Latency Sensitive	Yes	No	Yes
Synchronization	*Yes (GPS)	*Yes (access point)	No
Central Coordination	*No	*Yes	No
Topology Change	Fast	Slow	Medium
Payload Volume	Small	Small	Small
Traffic Pattern	Ad hoc	Convergecast	Ad hoc

* Not Always

2.4 Other Wireless MAC Protocols with Deterministic Performance

2.4.1 Self-Organizing Time Division Multiple Access (STDMA)

is a reservation-based channel access mechanism which relies on GPS equipped nodes rather than a central coordinator [16]. GPS time signal provides synchronization among the nodes and helps scheduling time slots, hence compensates lack of a central coordinator for providing deterministic channel access. Although STDMA is promising for VANETs with some commercial deployments, it is not suitable for CANETs where indoor affordable deployments are sought.

2.4.2 ADHOC-MAC

or equivalently *Reliable Reservation ALOHA* (RR-ALOHA) is a reservation based MAC protocol that is originally developed for inter-vehicle communications and relies on unconstrained power supply for each node in the network [17]. It considers limited coverage range of nodes and a true distributed mechanism in steady operation. It also does not rely on GPS or any similar central time synchronization method to facilitate TDMA channel access.

The basic idea in RR-ALOHA is that all of the nodes in a network are pre-programmed with the number of *time slots* in a *time frame*, and the duration of a time slot. When turned on, a node tries to reserve one of the time slots by chance. Collision can happen between competing nodes to reserve a specific time slot. When a node succeeds in reserving its own time slot, no other nodes in a two hop neighborhood may transmit during that interval. Therefore, this protocol completely eliminates *hidden terminal problem* and provides spatial reuse of the same time slot for far enough non-interfering neighbors. This is achieved in expense of obliging each node to gather information from its neighbors as far as two hops in order not to start any transmission during their reserved time slots. At the end, large MAC overhead imposed by the neighbor discovery messages significantly degrades performance of this protocol, especially in dense networks [18].

2.5 Viable MAC Solutions for Wireless PnP Control Systems

In summary, only the PA application domain of the IEEE802.15.4e and the EDCA mechanism of the IEEE802.11e can be applied to wireless PnP control systems. WTRP cannot appropriately handle the ring stability problem. Both LL networks of 802.15.4e and the HCCA mechanism of the IEEE802.11e require a central orchestrating node which cannot be provided when nodes have insufficient coverage range. Finally, RR-ALOHA was rejected due to large MAC overhead and other practical issues.

The bad news is that neither TSCH in the IEEE802.15.4e nor EDCA in the IEEE802.11e guarantee tight e2e latency requirements of PnP control systems. The former approach sacrifices the best achievable e2e latency by conservatively providing guarantees on latency bounds; and the latter cannot provide guarantees to its usually acceptable e2e latency. Therefore, TSCH in the IEEE802.15.4e is preferred when deterministic and fairly high latency is of interest. On the contrary, if low latency is crucial, EDCA in the IEEE802.11e should be employed with preparation to deal with jitter.

Last but not least, the IEEE802.15.4e has an obvious advantage over the IEEE802.11p when it comes to link reliability issues. Although the IEEE802.11p has 7 PHY channels, it does not employ them to provide a channel diversity mechanism as is done in the IEEE802.15.4e. Instead, it exploits those channels to prioritize packets in sake of collision avoidance. Moreover, within each channel, it uses the EDCA mechanism to further prioritize packets. The IEEE802.15.4e is superior because it uses different resources for different tasks, i.e. different PHY channels to provide channel diversity, and an individual PHY channel for accurate scheduling.

3 Selection of the Routing Protocol

If the coverage range of a wireless transmitting node is adequate to reach the receiving node, relaying messages among different hops is not needed, and hence no routing protocol is required just like single-segment wireline networks. However, an extensive coverage range has some drawbacks as well. It increases node density and heats up competition for gaining control of the channel. Thus, it makes the system prone to more interference and collisions at the MAC layer. Moreover, it requires higher transmission power which is constrained by governmental regulations. These two issues limit the coverage range of wireless nodes. In this paper, we have assumed that routing is an indispensable part of wireless PnP control systems.

The problem of routing consists of two tasks which are: 1) characterizing nodes with some sort of information or properties that can be used by a routing algorithm, and 2) finding a routing algorithm. Reference [19] offers an up-to-date chronological survey on routing protocols. In what follows, we have categorized those protocols in a different way to highlight different approaches in carrying out the above mentioned two tasks.

In terms of making a structure for the network, the following families of methods are identified in [19]:

1. No structure
2. Structuring based on geographical, relative or virtual coordinates
3. Structuring based on communication metrics

4. Application oriented structuring

Routing algorithms, on the other hand, can be generally categorized into three groups:

1. Routing based on a known sequence of addresses
2. Flooding based routing
3. Gradient based routing

A complete routing solution is a combination of a structuring method and a routing algorithm. The possible combinations are depicted in Fig. 10.4 and explained in details thereafter.

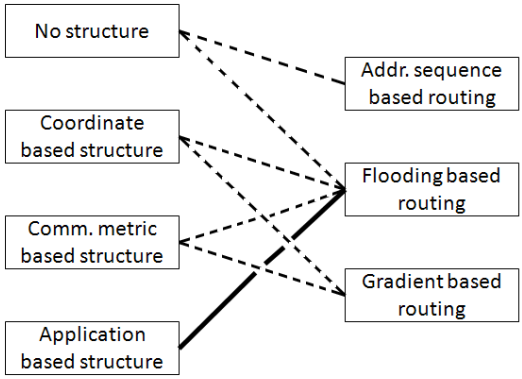

Figure 10.4: Possible combinations of structuring methods and routing algorithms to form a complete routing solution. The bold line shows the area of contribution of this paper.

3.1 Routing based on a known sequence of addresses

This is actually the most obvious way to route a packet to its destination. It is originally designed for multi-segment wireline networks. The packet contains a sequence of node addresses and it is forwarded from each node to the next one whose address is known. The problem lies in how to devise a working sequence of addresses in the source node.

A popular approach is to flood *route request* packets through the network before or when a transmission is to happen. The sequence address will be decided based on the replies to route request messages. There are a number of well-known routing protocols which use this principle including Dynamic Source Routing (DSR) [20], Ad-hoc On Demand Vector routing (AODV) [21], and Dynamic Mobile On-demand routing (DYMO) [22]. A simple and clear comparison between them is offered in [19].

3.2 Flooding-based routing

Chronologically, flooding is the first type of routing solutions that has appeared in wireless networks. Its name describes well how it works. When a node receives a packet, it checks whether it is the final recipient of the packet and if it has received the same packet

before. In pure flooding, i.e. flooding with no structure, a negative answer to both of these questions results in retransmission of the packet. In this way, a packet is flooded throughout the entire network and will be received by every node in a connected segment.

Pure flooding has several advantages. First of all, it is a reliable routing method. If there is a good path from source to destination, the packet will eventually find its way and arrives at the final destination as soon as possible. Pure flooding is also the most robust routing method against topological modifications because there is no structure in it that needs to be updated after a topological change. But unfortunately, it is painfully resource consuming. Too many healthy retransmissions are required before a packet is faded from the network. This will not only consume much energy, which is valuable especially in battery-operated nodes, but also will increase conflicts and collision chance at the MAC layer.

Several remedies are proposed to reduce the number of unnecessary retransmissions in flooding-based routing. Some assume that a limited number of retransmissions are enough to reach the destination and do not propagate the packet any further. Such a number is derived either probabilistically [23] or by considering the worst case [24]. Others impose a structure to the network. In coordinate-based structuring, information either about the physical location or the virtual location of nodes in a system hierarchy is used to form clusters and flood packets only among cluster members [25]. Communication metrics might also be involved in setting a structure for the network e.g. by omitting the nodes that have little residual energy or by blacklisting the paths which are not reliable enough [26]. The last method, i.e. application-based structuring proposes exploiting similarity of data content and basically dropping out redundant data [27].

3.3 Gradient-based routing

Gradient based routing utilizes the structure of the network to relay packets towards one or a few number of data sink nodes. This approach implies assigning a scalar value – per sink – to all nodes, representing their *distance* from that sink. Distance, could be any kind of routing metric [28]. A route is then selected between the transmitter and the receiver nodes such that the shortest distance is taken between the nodes.

Most of the gradient routing methods are coordinate based. They require all of the nodes to know either their geographical, relative, or virtual locations. It is realized by different methods including: pre-programming of fixed nodes, use of GPS [29], use of anchor nodes [30], or calculation of virtual coordinates [31]. After assigning location of nodes by any of these four methods, the coordinates of the destination is stored in each packet before leaving the source node. Then it can be forwarded towards the destination either by a greedy approach or a combination of greedy and circumvolutory modes to escape network holes and guarantee packet delivery [32].

Other gradient-based routing methods utilize various communication metrics in order to define and assign nodes ranks. A comprehensive example is the in-progress routing standard by Internet Engineering Task Force (IETF), namely Routing Protocol Over Low power and Lossy networks (RPL) [33]. IETF RPL is the pioneer among routing standards since standardization activities used to be limited to PHY and MAC layers. It is based on the ratified *tree* structure which may utilize various communication metrics in order to prioritize links or nodes in an accumulative or individual way [34]. IETF RPL, like any other gradient-based routing method, is mainly designed for MP2P traffic. However,

other traffic patterns such as point to multi-point (P2MP) and P2P are also supported in the IETF RPL. This means that P2P and P2MP are exceptions, only established after a node's request, and maintained temporarily [35].

Unlike the previous section, which was ended with recommendations on selection of the most viable MAC solutions, the same task in this section is postponed until some assumptions on data traffic pattern and the network structure is clarified in the next section.

4 The Routing Solution

In a wireless PnP control system, there are three kinds of nodes: actuators, controllers, and sensors. All of them should have the capability to act both as a data *sink* and as a data *source*, described as follows.

- A Sensor is regularly a data source to send sensory data towards relevant controllers, typically once per control cycle time, e.g. 100 ms.
- A sensor sporadically acts as a data sink to receive configuration data from its associated controllers.
- An actuator is regularly a data sink which receives commands from the associated controller at each control cycle time and implements them.
- An actuator sporadically acts as a data source to report failures.
- Controllers should send and receive data in each control cycle time interval. They gather data from sensors at the beginning of a typical cycle time interval, and send commands to the actuators at the end of the interval.

Unlike gradient-based routing protocols, which assume a Multi-point to Point (MP2P) traffic with controllers as the sink nodes [34], the above items suggest a P2P traffic pattern, especially because neither sensors nor actuators outnumber each other in comprising PnP control systems. This topological assumption makes gradient-based routing incompatible with wireless PnP control systems. Therefore, we decided to focus on flooding-based routing.

Fig. 10.2 in the beginning of the paper depicts the key assumptions of topology design, described in the following:

1. There could be a large number of controller units ($\Lambda_i, i = 1, ..., n$).
2. Each controller and its associated sensors and actuators form a set, called a *cluster* henceforth. Clusters are identified by their unique controller unit tag ($\Lambda_i, i = 1...n$).
3. There are as many clusters as controller nodes which are called *cluster heads*.
4. Each packet contains a cluster *association* field. In general, a packet might be tied to one or more clusters. It is also possible that a packet is not associated with any cluster.

5. A sensor might be a *member* of multiple clusters, meaning that its generated data could be associated to more than one cluster head. In other words, a packet that is generated at a sensor node might be reported to more than one controller.

6. An actuator could be a member of at most one cluster, meaning that it may not receive commands from more than one controller.

7. Data packets to/from members of a cluster should be sent from/to the cluster head. In other words, a controller node is either the source or the final destination in every transmission path.

4.1 An application based routing metric

According to the state of the art in [28], none of the routing metrics in the literature has made a direct connection between the top layer application and the networking protocol so far. We intend to fill this gap in this section by proposing such a routing metric as the main contribution of the paper.

Our routing metric is actually a network structuring method based on the control application. This method works by classifying the nodes into different clusters based on their relevance to different control loops. See Fig. 10.2 in the beginning of the paper. To this end, we rely on a theoretical result in [3]. It offers three stochastic correlation-based measures that indicate *usefulness* of incorporating a new sensor y_a or a new actuator u_a in a present system model, which is considered as follows:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu_p(k) + \omega(k) \\ y_p(k) &= Cx(k) + Du_p(k) + v(k) \end{aligned} \tag{10.1}$$

In (10.1), x, u_p and y_p stand for states, present inputs and present outputs of the discrete time LTI system model with the input disturbance $\omega(t)$ and the measurement noise $v(t)$.

The three mentioned measures are employed as application-based routing metrics in order to extend the concept of *distance*. Consequently, membership in a specific cluster Λ_j is granted to a node if that node is closer to the cluster head than a specific threshold value d_j.

4.1.1 Addition of a new sensor

Addition of a new sensor is illustrated in Fig. 10.5. For a newly added sensor node, the following two complementary measures are proposed.

$$d^2_{U_p,y_a} = \frac{E[y_a(k) - \hat{y}_a(k|U_p^{k-1})]^2}{E[y_a(k) - E(y_a(k))]^2} \tag{10.2}$$

$$d^2_{U_pY_pY_a,y_a} = \frac{E[y_a(k) - \hat{y}_a(k|U_p^{k-1}, Y_p^{k-1}, Y_a^{k-1})]^2}{E[y_a(k) - E(y_a(k))]^2} \tag{10.3}$$

in which d^2 represents the correlation based distance and varies between 0 and 1. Subscripts $(.)_p$ and $(.)_a$ refer to present model and added device, respectively. $y(k)$ and $u(k)$ mean individual samples of a sensor's data and an actuator's command at time k, while Y^k and U^k indicate the set of all samples from the beginning up to and including time k.

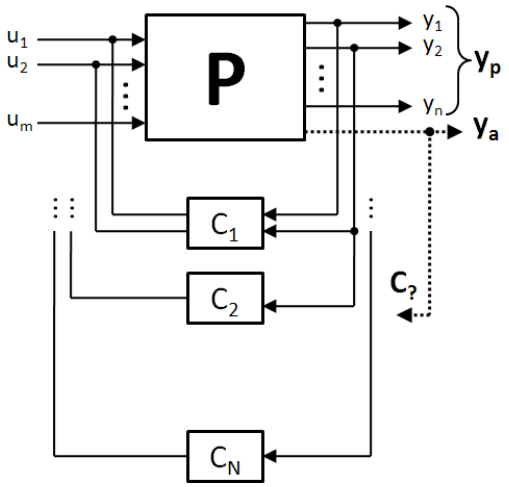

Figure 10.5: Block diagram of a PnP control system: The case of addition of a new sensor that should be connected to the relevant controllers

$E(.)$ stands for expected value operator over a finite number (N) of data samples which are treated as random variables [3]. N is pre-defined in the sensor node. Superscript $\hat{(.)}$ stands for the least-squares estimation based on the available model.

With respect to the above mentioned definitions, interpretation of (10.2) and (10.3) is given in the following paragraph, assuming that: 1) the model of the present system is discrete-time linear time-invariant and 2) the new sensor provides sufficiently exciting data, and 3) a consistent un-biased least-squares estimation is given when $N \rightarrow \infty$.

The denominator in (10.2) and (10.3) is the variance of the data gathered by the new sensor, i.e. y_a. The numerator in (10.2) indicates how predictable the current $y_a(k)$ is if the commands of all actuators are known in the previous samples. If, according to the present model, none of the $k-1$ samples of all of the actuators have any tangible effect on the k^{th} sample of y_a, the following equation holds true.

$$E[y_a(k)|U_p^{k-1}] = E[y_a(k)] \tag{10.4}$$

Furthermore, if an unbiased estimation is assumed, we have $\hat{y}_a(k|U_p^{k-1}) = E[y_a(k)|U_p^{k-1}]$ which in combination with (10.4) results in the following expression:

$$\hat{y}_a(k|U_p^{k-1}) = E[y_a(k)] \tag{10.5}$$

Equation (10.5) means that the conditional least squares estimation of y_a is equal to its actual expected value. Therefore, the present model is good enough and the new measurement does not add any value to it. In this situation, $d^2_{U_p,y_a} = 1$, which should be read as: the new node is too far from the cluster head and cannot become a member, that is it is irrelevant to the control loop in question.

Equation (10.2) measures how much the additional sensor is affected by the present actuators in open loop. Nevertheless, this measure only reveals linear correlation. To look for nonlinear correlations, (10.2) should be modified according to the specific nonlinearity

we are looking for. This is the easy step, but the difficult part is to perform nonlinear online incremental system identification to find $\hat{y}_a$. We do not consider this case in this paper.

The numerator in (10.3) measures how much the additional output could be controlled by the present actuators in closed loop. The interpretation is similar to (10.2), but this time the data from all of the sensors, including the new one, are also used in the least squares estimation, hence making it a more computationally intensive problem. Either (10.2) or (10.3) could be used in a given setting. Exploiting (10.2) is recommended in cases where y_a cannot be controlled independently of y_p [3].

4.1.2 Addition of a new actuator

When a new actuator is added, the following measure is proposed.

$$d^2_{U_a y_p | U_p Y_p} = \frac{E[y_p(k) - \hat{y}_p(k|U_p^{k-1}, Y_p^{k-1}, U_a^{k-1})]^2}{E[y_p(k) - \hat{y}_p(k|U_p^{k-1}, Y_p^{k-1})]^2} \tag{10.6}$$

Equation (10.6) measure how much influence the additional actuator has on the present sensors in closed loop. If U_a^{k-1} does not have any effect on improving prediction of $\hat{y}_p$, then the prediction errors in the numerator and denominator will look alike and d will get its maximum value ≈ 1 On the other hand, if U_a^{k-1} is useful such that the prediction error in numerator is much less than that in denominator, then we have: $d \to 0$. Equation (10.6) should be interpreted similar to the previous measures with similar concerns. The same assumptions hold for a consistent estimation of $\hat{y}_p$. In practice, to provide a sufficiently exciting control signal, the actuator has to be driven by an external signal.

4.2 Cluster Formation

Here, we give a high level description of cluster formation. Assume that the controller nodes ($\Lambda_i, i = 1, ..., n$) are deployed as cluster heads. In a realistic scenario, each cluster head has a built-in model of the subsystem it is supposed to control. All of the initially deployed sensor and actuator nodes are already bound to their controllers. In other words, in the network setup phase, all of the nodes are aware of their cluster membership. As a result, sensor nodes immediately start to function in their normal operating mode. Actuator nodes should start working in a safe mode and wait until they receive commands from the cluster head, i.e. the controller unit. The cluster head starts sending commands to the actuator as soon as it can devise the commands based on received sensor packets and the pre-programmed plant model. In both above cases, the packets flood their pertinent clusters.

Later on, when a new sensor pops up, it does not initially belong to any cluster and it is in transient operating mode. Thus, it publishes data in *roaming packets*. Fig. 10.6 illustrates an example when a new sensor is placed somewhere in a network such that its roaming packets could reach two cluster heads, e.g. Λ_1 and Λ_2. Based on the above algorithm, both cluster heads start calculating (10.2), (10.3), or both. Fig. 10.6 shows how the correlation-based distance measures (10.2) or (10.3) might develop over time in Λ_1 and Λ_2. It is assumed that the new sensor generates 1200 roaming packets.

After collecting sufficient samples at the cluster heads, each cluster head decides whether the new sensor should become a member of their cluster or not. In Fig. 10.6,

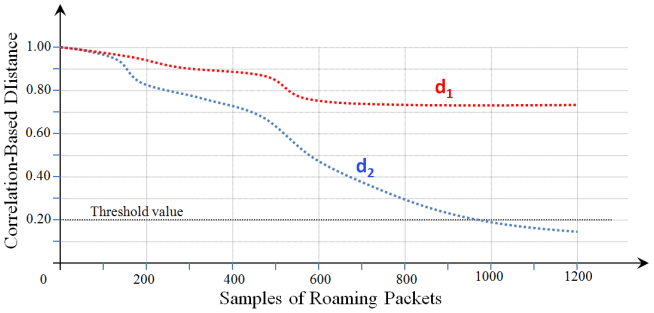

Figure 10.6: A new sensor is turned on in a network with several clusters and has started a PnP start-up procedure. Data from the sensor node has reached two cluster heads Λ_1 and Λ_2 during the joining process. This graph illustrates development of the usefulness measure (10.2) in both cluster heads.

Λ_2 concludes that the new sensor is relevant, before receiving 1000 samples, but Λ_1 finds the new sensor irrelevant to the control performance of its internal model. Λ_1 stops any further action when roaming packets are discontinued, but Λ_2 continues by sending a *join* request to the new sensor as soon as the threshold value is passed. After sending all roaming packets, equal to 1200 in our example, the new sensor applies the received join requests. Join requests arrive at the sensor node asynchronously. Therefore, the node should continually accept join requests, at least until a pre-defined time. If no join request is received, the sensor starts another round of generating roaming packets.

The above procedure is slightly different for a new actuator. Each newly turned on actuator applies a pre-specified control sequence for the purpose of sufficiently exciting the plant and creating measureable outcomes. Simultaneously, it publishes roaming packets which contain current value of the actuator output. The cluster heads which receive these packets, use the value of the actuator output in evaluating (10.6) similar to what was shown in Fig. 10.6. When the roaming packets are discontinued – meaning that the actuator is waiting for the decision – each cluster head returns the calculated *usefulness measure* to the actuator. The actuator waits for a certain time to receive evaluation results from all potentially involved cluster heads. Then it compares the received values and selects the cluster that has returned the largest one. If at least one evaluation result is received, the actuator chooses its own cluster and sends a join request to that cluster. If no evaluated usefulness measure is received, the above procedure starts from the beginning.

Note that, when a new sensor is added, it plausibly receives individual join requests from multiple controllers. But when a new actuator is added, it sends the join request to a single controller.

Last but not least, the effect of e2e latency should be studied on the results in Fig. 10.6. It is expected that by imposing latency on packets during the identification process, parameters of the identified model experience difficulties in convergence. That will result in development of inaccurate models which significantly affect the control performance. This phenomenon is illustrated in a simulation example in the next section.

5 Simulation results

In this section, we evaluate effects of the e2e latency and jitter on our data-centric clustering method via a simulation example. The simple version of this example without considering communication constraints is already presented in [3]. Our purpose is to find out how the system model and the network structure will develop if a new sensor joins the network in presence of latency and jitter. The original system model is:

$$\begin{bmatrix} x_1(t+1) \\ x_2(t+1) \end{bmatrix} = \begin{bmatrix} 0.9048 & -0.0090 \\ 0.0090 & 0.9048 \end{bmatrix} . \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0.0952 & 0.0947 \\ 0.00047 & 0.0956 \end{bmatrix} . u_p(t) + \begin{bmatrix} w_1(t) \\ w_2(t) \end{bmatrix} \tag{10.7}$$

$$y_p(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} . \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 & 0.1 \end{bmatrix} . u_p(t) + v(t) \tag{10.8}$$

with noise correlations:

$$R_w = \begin{bmatrix} 0.0100 & 0.0050 \\ 0.0050 & 0.0100 \end{bmatrix}, R_v = 0.01, R_{wv} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{10.9}$$

The original system is intentionally chosen such that $x_2(t)$ has poor observability. This is implied by the large condition number (≈ 201) of the observability matrix. In the simulation example in [3], a fast LQ state feedback controller (10.10) is used in combination with a kalman filter whose gain and innovation covariance are given in (10.11).

$$u(t) = - \begin{bmatrix} 2.3953 & -1.2122 \\ 1.2138 & 2.4299 \end{bmatrix} . \hat{x}(t) \tag{10.10}$$

$$K_p = \begin{bmatrix} 0.5381 \\ 0.2597 \end{bmatrix}, \quad R_{ep} = 0.0248 \tag{10.11}$$

It is also illustrated in [3], how a new output measurement signal $y_2(t)$, as described in (10.12), will contribute to improvement of: 1) observability of the second state, and 2) the total cost of the unaltered LQ controller by just updating the kalman filter.

$$y_2(t) = \begin{bmatrix} 1 & 1 \end{bmatrix} . \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0.1 & 0.1 \end{bmatrix} . \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix} + v_a(t) \tag{10.12}$$

The block diagram of the system is shown in Fig. 10.7. It is a special case of the general problem setup in Fig. 10.5, in which there is only one controller and we are interested to see if the added remote sensor will improve control performance. If there were more than one controller, the effect on control performance of each controller would imply whether the new sensor should stay connected to that controller, and hence become a member of its cluster.

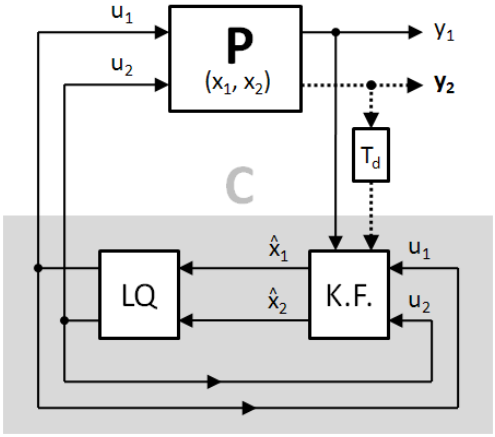

Figure 10.7: A simple PnP control system which benefits from addition of a new sensor whose data is delayed T_d due to network latency

Assuming the state and control input cost functions are defined as in (10.13) and (10.14), and taking no communication constraints into account, i.e. $T_d = 0$, improvements of the incremental model compared to the original model are shown in Tab. 10.2 for the sake of completeness.

$$\begin{aligned} J_x &= \frac{1}{N}\sum_{t=1}^{N} x(t)^T Q_x x(t) \\ &= \frac{1}{N}\sum_{t=1}^{N} x_1(t)^2 + x_2(t)^2 \end{aligned} \tag{10.13}$$

$$\begin{aligned} J_u &= \frac{1}{N}\sum_{t=1}^{N} u(t)^T Q_u u(t) \\ &= 0.05\frac{1}{N}\sum_{t=1}^{N} u_1(t)^2 + u_2(t)^2 \end{aligned} \tag{10.14}$$

Table 10.2: Modification of the state cost, the control cost, and the total cost of the updated system compared to the original system [3]

	J_x	J_u	$J = J_x + J_u$
Updated/Original	0.5851	1.6265	0.6340

In this paper, we have extended the above mentioned example by considering the original model as a lumped model without any communication constraint, but the added

sensor as a remote sensor, whose data has to be transferred via a network. The network will induce a latency (T_d) on the newly added measured data as expressed in (10.15).

$$y_a(t) = y_2(t - T_d) \tag{10.15}$$

Assuming stability of the routes through the network, which is mainly provided by the channel hopping and channel agility mechanisms recommended in Section II, the induced latency dynamics is affected only by the MAC protocol. Since either the IEEE802.15.4e or the IEEE802.11e were already chosen as the MAC layer solution, two scenarios were possible. In the first one, a relatively large latency without jitter was imposed to the data from the added sensor. It is meant to be understood as the abstracted effect of choosing TSCH in the PA application domain of the IEEE802.15.4e. In the second scenario, associated with the EDCA mode of the IEEE802.11e, relatively small latencies were applied which were subject to non-zero jitter with uniformly distributed values.

Fig. 10.8 shows simulation results for the first scenario. Among the cost functions in Tab. 10.2, only the state cost is of interest because the control cost is not improved even in the original example without any latency. Actually, any improvement in the state cost is achieved by employing larger control signals. The total cost is also a summation which can easily be modified by changing Q_x and Q_u in (10.14) and (10.13). That is why only the state cost is investigated and illustrated in Fig. 10.8.

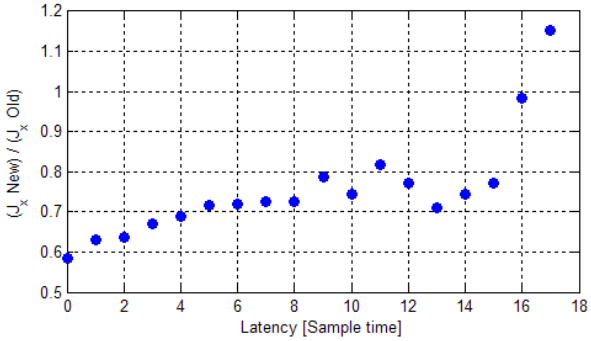

Figure 10.8: Influence of jitter-free latency on relative state cost

As Fig. 10.8 shows, control performance degrades as latency increases. This continues until a certain amount of latency after which, the identification process fails and the use of an incorrect model results in very large oscillations, i.e. stochastic behavior, in the state cost. This means that the parameters of the updated model are not converged anymore.

In this example, the largest admissible latency is 9 times larger than one sample time. It can be seen that with such a constant latency, improvement on the state cost from 41.5% for a latency-free network to 21.2% for a network with 100% reliable links and completely deterministic MAC protocol as in TSCH mode of the IEEE802.15.4e.

Fig. 10.9 adds the effect of latency variation on the relative state cost. Each curve in this figure represents a stochastic specification for jitter. The simulation is done on the assumption that each roaming packet from the newly joined sensor is subject to a

random latency which is a multiple of the sampling rate and is uniformly distributed in the interval $[Latency_{min}, Latency_{min} + 2 \times Jitter]$. Each curve is produced by increasing $Latency_{min}$ until numerical simulations become invalid, as explained above.

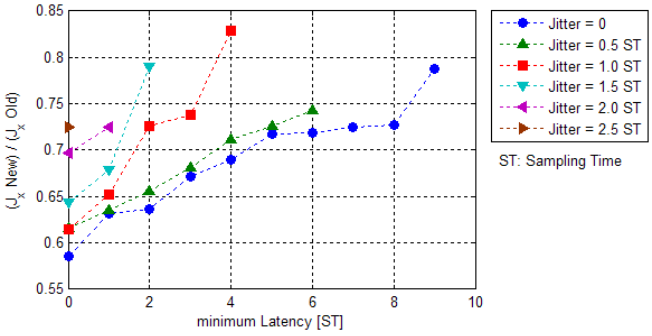

Figure 10.9: Influence of latency and jitter on relative state cost

Assuming that the average latency in all of the above cases equals to $Latency_{min} + Jitter$, the bar plot in Fig. 10.10 shows how jitter reduces the maximum tolerable average latency. Each bar in this figure is plotted based on the end point of individual curves in Fig. 10.9. It can be seen how deterministic behavior contributes to pushing the limits of the tolerable latency forward.

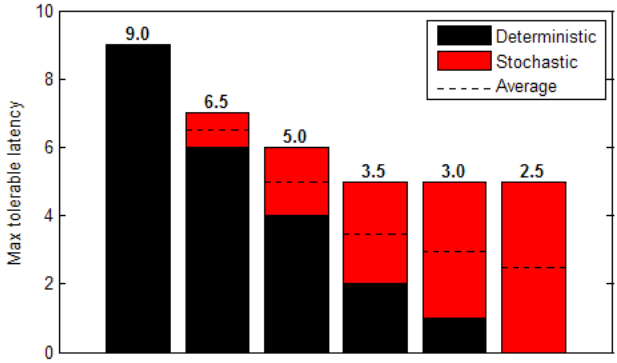

Figure 10.10: Maximum tolerable average latency

The simulation results confirm that wireless PnP control systems benefit most from completely deterministic solutions at the MAC layer. They also show how a control oriented scalar measure can act as a routing metric and contribute to clustering of the nodes of a network in a flooding-based routing algorithm. The proposed routing metric makes use of model-based correlation estimation between the new nodes and the existing model of the system.

6 Conclusion

This paper studied the communication framework of a wireless PnP control system, its inherent problems in protocol design and possible solutions in MAC and NTW layers. Recommendations on choosing viable MAC and routing protocols are given in Section II.E and Section III.D, respectively. The main contribution of the paper is to make a connection between the control topology and the networking topology by introducing a new routing metric which is truly application dependent. The routing metric is developed by the control application and is used in classifying the nodes of a wireless network into different clusters. Functionality of the proposed routing metric is illustrated in a simulation example by considering assumptions on the MAC and routing protocols in accordance with the recommendations in Section II.E and Section III.D.

Acknowlegement

This research was supported by the Plug and Play Process Control (P^3C) project. P^3C was partly sponsored by the Danish Research Council for Technology and Science Production.

References

[1] J. Stoustrup, "Plug & play control: Control technology towards new challenges," *European Journal of Control*, vol. 15, no. 3-4, pp. 311–330, 2009.

[2] J. Bendtsen, K. Trangbaek, and J. Stoustrup, "Plug-and-play controlmodifying control systems online," *IEEE Transactions on Control Systems Technology*, vol. PP, no. 99, pp. 1–15, November 2011.

[3] T. Knudsen, J. Bendtsen, and K. Trangbaek, "Awareness and its use in incremental data driven modelling for Plug and Play Process Control," *European Journal of Control*, vol. 18, no. 1, pp. 24–37, February 2012.

[4] T. Knudsen, "Awareness and its use in plug and play process control," in *European Control Conference*, Budapest, Hungary, August 2009.

[5] J. Bendtsen, K. Trangbaek, and J. Stoustrup, "Plug-and-play process control: Improving control performance through sensor addition and pre-filtering," in *Proceedings of the 17th IFAC World Congress*, July 2008.

[6] IETF ROLL: Industrial routing requirements. [Online]. Available: http://tools.ietf.org/html/rfc5673

[7] A. Willig, K. Matheus, and A. Wolisz, "Wireless technology in industrial networks," *Proceedings of the IEEE*, vol. 93, no. 6, pp. 1130–1151, May 2005.

[8] A. Willig and A. Wolisz, "Ring stability of the PROFIBUS token-passing protocol over error-prone links," *IEEE Transactions on Industrial Electronics*, vol. 48, no. 5, pp. 1025–1033, October 2001.

[9] A. Willig, "Analysis of the PROFIBUS token passing protocol over wireless links," in *IEEE International Symposium on Industrial Electronics*, vol. 1, 2002, pp. 56–60.

[10] M. Ergen, D. Lee, R. Sengupta, and P. Varaiya, "WTRP – Wireless Token Ring Protocol," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 6, pp. 1863–1881, November 2004.

[11] A. Bachir, M. Dohler, T. Watteyne, and K. K. Leung, "MAC essentials for wireless sensor networks," *IEEE Communications Surveys & Tutorials*, vol. 12, no. 2, pp. 222–248, November 2010.

[12] L. Winkel, Z. Sahinoglu, and L. Li, *IEEE Std 802.15.4e-D0.01*, IEEE P802.15 Working Group for Wireless Personal Area Networks Draft, Rev. 06, March 2010.

[13] K. S. J. Pister and L. Doherty, "TSMP: time synchronized mesh protocol," in *Proceedings of the IASTED International Symposium on Distributed Sensor Networks*, November 2008.

[14] T. Watteyne, A. Mehta, and K. Pister, "Reliability through frequency diversity: why channel hopping makes sense," in *Proceedings of the 6th ACM symposium on Performance evaluation of wireless ad hoc, sensor, and ubiquitous networks*, 2009.

[15] A. Willig, "Recent and emerging topics in wireless industrial communications: A selection," *IEEE Transactions on Industrial Informatics*, vol. 4, no. 2, pp. 102–124, May 2008.

[16] H. Lans, U.S. Patent 5,506,587, 1996.

[17] F. Borgonovo, A. Capone, M. Cesana, and L. Fratta, "ADHOC MAC: a new MAC architecture for ad hoc networks providing efficient and reliable point-to-point and broadcast services," *ACM Wireless Networks*, vol. 10, no. 4, pp. 359–366, July 2004.

[18] ——, "RR-ALOHA, a Reliable R-ALOHA broadcast channel for ad-hoc inter-vehicle communication networks," in *Med-Hoc-Net*, 2002.

[19] T. Watteyne, A. Molinaro, M. G. Richichi, and M. Dohler, "From MANET to IETF ROLL standardization: A paradigm shift in WSN routing protocols," *IEEE Communications Surveys & Tutorials*, vol. 13, no. 4, pp. 688–707, Fourth Quarter 2011.

[20] D. B. Johnson, D. A. Maltz, and J. Broch, *DSR: The Dynamic Source Routing Protocol for Multi-Hop Wireless Ad Hoc Networks*. Addison-Wesley, 2001, ch. 5, pp. 139–172.

[21] C. Perkins, E. Belding-Royer, and S. Das, *Ad hoc On-Demand Distance Vector (AODV) Routing*, Network Working Group Std., July 2003.

[22] I. Chakeres and C. Perkins, *Dynamic MANET On-demand (DYMO) Routing*, Mobile Ad hoc Networks Working Group draft, Rev. 21, July 2010.

[23] M. Hamdi, N. Essaddi, and N. Boudriga, "Energy-efficient routing in wireless sensor networks using probabilistic strategies," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Las Vegas, NV, USA, April 2008, pp. 2567–2572.

[24] T. Clausen and P. Jacquet, *Optimized Link State Routing Protocol (OLSR)*, Network Working Group Std. RFC 3626, October 2003.

[25] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 660–670, October 2002.

[26] R. Shah and J. Rabaey, "Energy aware routing for low energy ad hoc sensor networks," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Orlando, FL, USA, March 2002, pp. 17–21.

[27] J. Kulik, W. Heinzelman, and H. Balakrishnan, "Negotiation-based protocols for disseminating information in wireless sensor networks," *Wireless Networks*, vol. 8, no. 2, pp. 169–185, March-May 2002.

[28] IETF ROLL: routing metrics used for path calculation in low power and lossy networks. [Online]. Available: http://tools.ietf.org/wg/roll/draft-ietf-roll-routing-metrics/

[29] T. Imielinski and J. C. Navas, "Gps-based geographic addressing, routing, and resource discovery," *Communications of the ACM*, vol. 42, no. 4, pp. 86–92, 1999.

[30] L. Blazevic, J. L. Boudec, and S. Giordano, "A location-based routing method for mobile ad hoc networks," *IEEE Transactions on Mobile Computing*, vol. 4, no. 2, pp. 97–110, March-April 2005.

[31] M. Chen, X. Wang, V. C. M. Leung, and Y. Yuan, "Virtual coordinates based routing in wireless sensor networks," *Sensor Letters*, vol. 4, no. 3, pp. 325–330, September 2006.

[32] H. Frey and I. Stojmenovic, "On delivery guarantees of face and combined greedy-face routing in ad hoc and sensor networks," in *Proceedings of the 12th annual international conference on Mobile computing and networking*. ACM, 2006.

[33] [Online]. Available: http://tools.ietf.org/wg/roll/

[34] IETF ROLL: IPv6 Routing Protocol, draft standard. [Online]. Available: http://tools.ietf.org/wg/roll/draft-ietf-roll-rpl/

[35] IETF ROLL: P2P traffic support. [Online]. Available: http://tools.ietf.org/wg/roll/draft-ietf-roll-p2p-rpl/